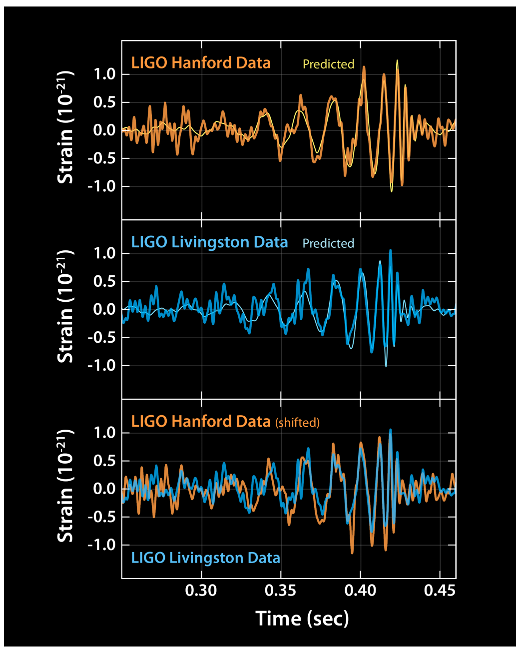
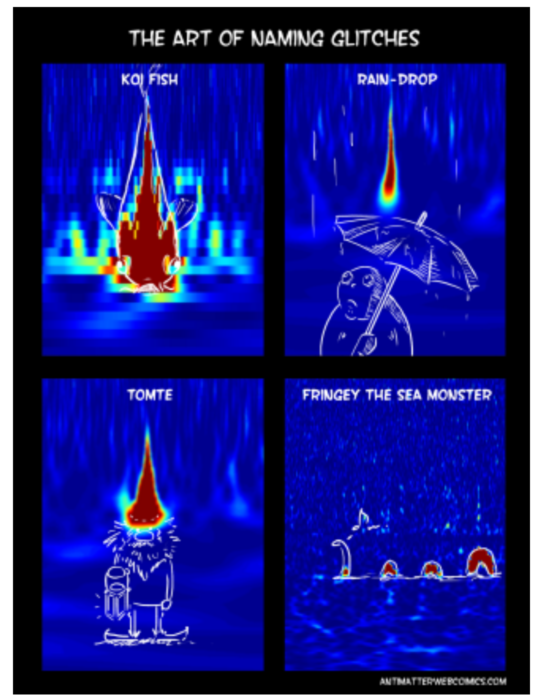
Searching for Gravitational Waves with LIGO: the role of blind injections and blind analyses

Alan J Weinstein
LIGO Laboratory, Caltech
LIGO Scientific Collaboration



LIGO Laboratory, Caltech

KIPAC Blind Analysis Workshop
March 13, 2017

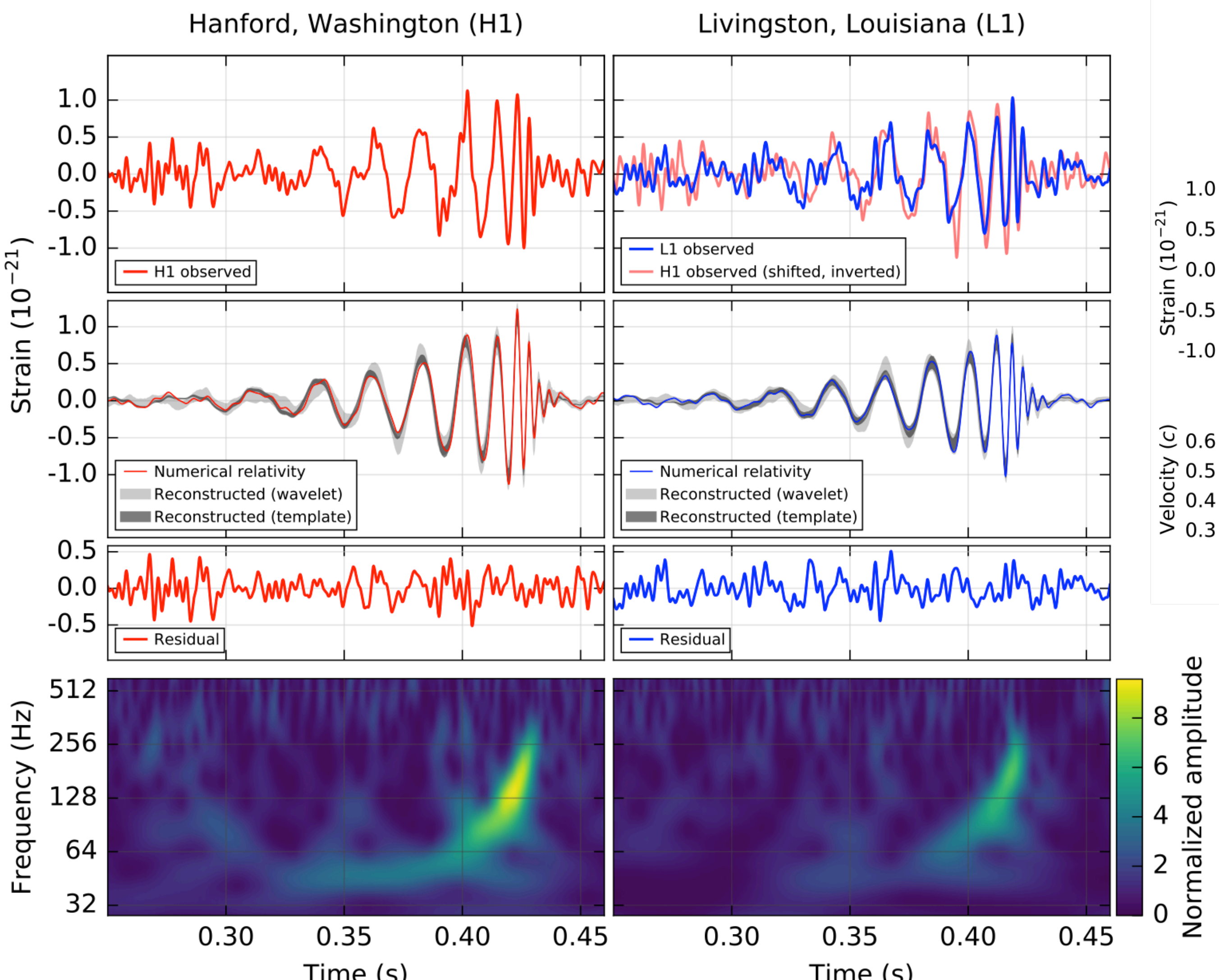


N. Kijbunchoo, LIGO

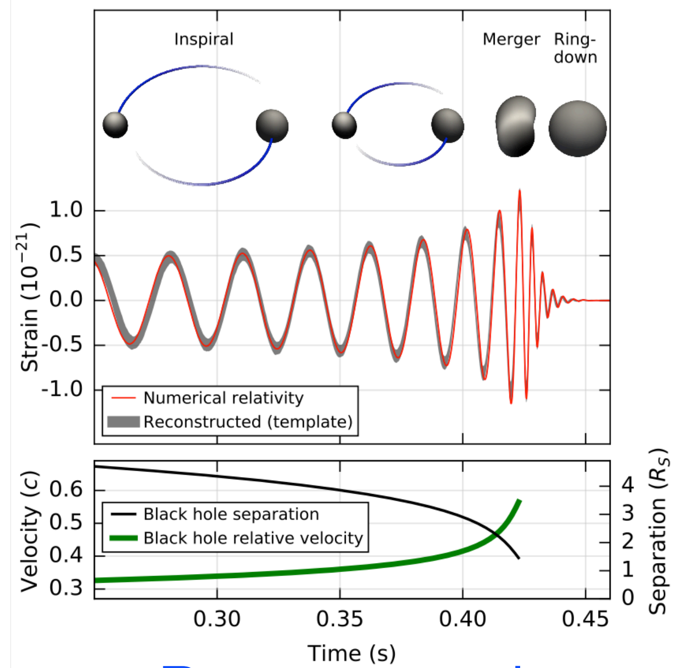
LIGO-G1700539

Discovery: GW150914

Phys. Rev. Lett. 116, 061102 – Published 11 February 2016

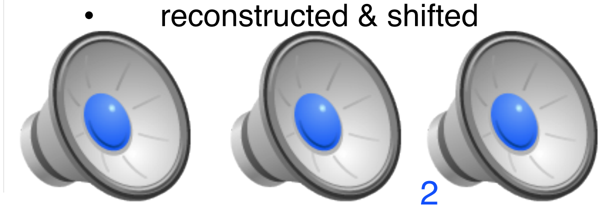


Whitened and band-passed [40-300] Hz



Reconstructed (no whitening)

Audio:
- filtered data
- freq-shifted data
- reconstructed & shifted

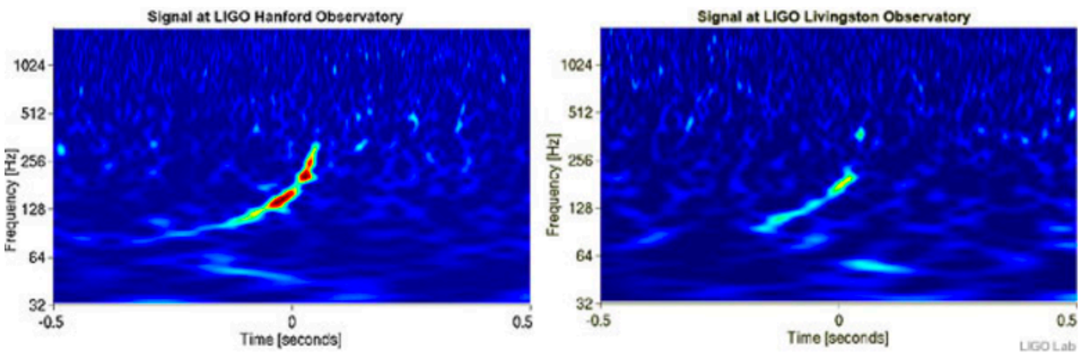


LIGO's FIRST "discovery": the BIG DOG

THE SIGNAL:

A rather strong signal was observed on September 16, 2010, within a minute or so of its apparent arrival at the detectors. The scientists on duty at the detector sites immediately recognized the tell-tale chirp signal expected from the merger of two black holes and/or neutron stars, and sprang into action. They knew that it could be a blind injection, but they also knew to act like it was the real thing.

The event was beautifully consistent with the expected signal from such a merger. The figures below show the strength of the signal (redder colors indicate more signal power) in time (horizontal axis) and frequency (vertical axis). The signal sweeps upwards in frequency ("chirp") as the stars spiral into one another, approaching merger. The first plot is what was seen in the LIGO Hanford detector, and the second is what was seen at the same time in the LIGO Livingston detector. Despite apparent differences, the two signals are completely consistent with one another. The dark and light blue regions are typical of fluctuating noise in the detectors.

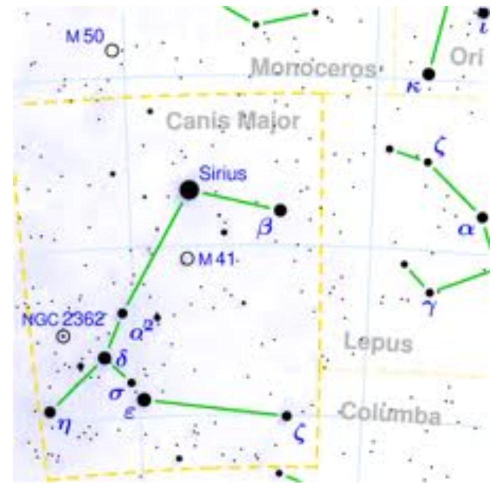


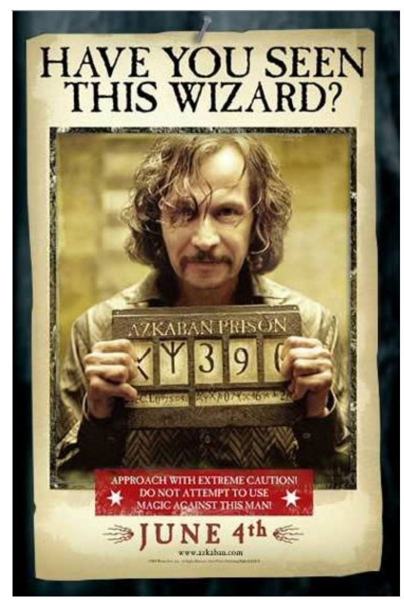
The loudness of the signal was consistent with it coming from a galaxy at a distance between 60 and 180 million light-years from ours.

The detector network is capable of locating the source in the sky only crudely; it seemed to be coming from the constellation Canis Major (the "Big Dog") in the southern hemisphere (the event was dubbed "the Big Dog" shortly thereafter). They sent alerts to partners operating robotic optical telescopes in the southern hemisphere (ROTSE, TAROT, Skymapper, Zadko) and the Swift X-ray space telescope, all of which took images of the sky on that and/or subsequent days in the hope of capturing an optical or X-ray "afterglow".

IS IT REAL?

In the subsequent days and weeks, numerous teams of scientists tried to get definitive answers to many questions. The event was seen strongly in the two LIGO detectors, less strongly in the Virgo detector, and nothing was seen in the less sensitive GEO detector. Could the event be explained as an accidental



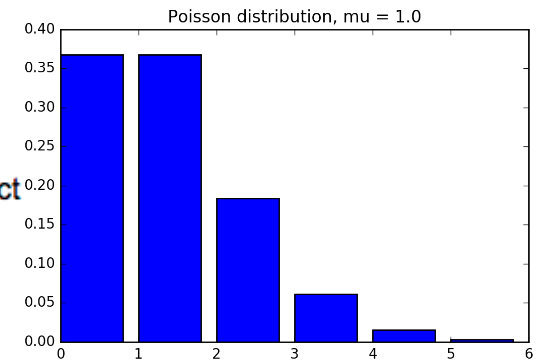


Blind injections

BLIND INJECTIONS:

The LIGO Scientific Collaboration and the Virgo Collaboration conducted their latest joint observation run (using the LIGO Hanford, LIGO Livingston, Virgo and GEO 600 detectors) from July, 2009 through October 2010, and are jointly searching through the resulting data for gravitational wave signals standing above the detector noise levels. To make sure they get it right, they train and test their search procedures with many simulated signals that are injected into the detectors, or directly into the data streams. The data analysts agreed in advance to a "blind" test: a few carefully-selected members of the collaborations would secretly inject some (zero, one, or maybe more) signals into the data without telling anyone. The secret goes into a "Blind Injection Envelope", to be opened when the searches are complete. Such a "mock data challenge" has the potential to stress-test the full procedure and uncover problems that could not be found in other ways.

The outcomes from previous blind injection exercises were reported in 2010, in this publication and this publication.



- We have an event! Nobel Prize!! Stay up all night elevating the event!
- Nah, it's probably a blind injection – kill the event, get some sleep.
- So, what'll it be? Nobel prize, or a good night's sleep?
- Argh!!! Confirmation bias!
- This was the genius of the fake signal injection: Whatever the prior belief of an individual scientist might be, it gave him or her reason to doubt it.

We wrote a paper …
(Observation / evidence / discovery ...)

Evidence for the Direct Detection of Gravitational Waves from a Black Hole Binary Coalescence

The LIGO Scientific Collaboration[1] and The Virgo Collaboration[2]

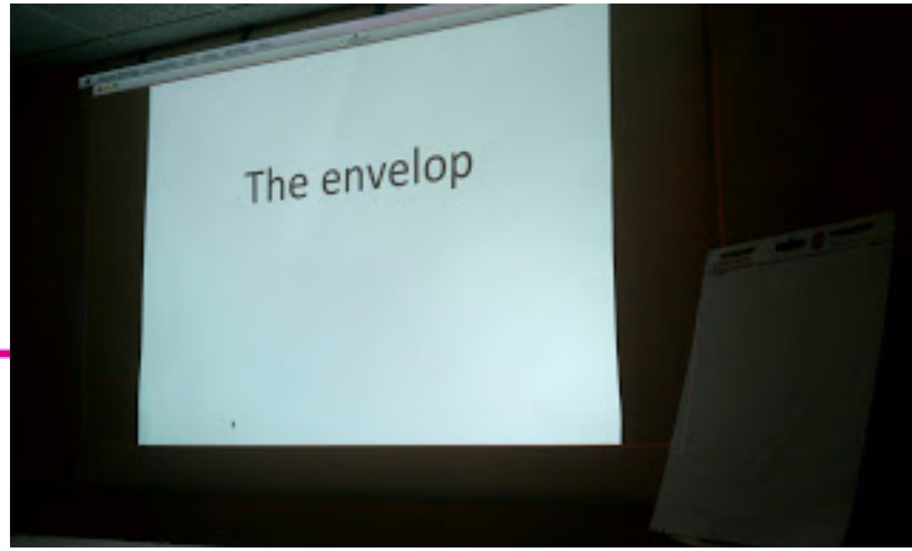
[1] *The LSC*
[2] *Virgo*

(RCS Id: detection.tex,v 1.81 2011/03/09 19:03:31 ajw Exp ; compiled 9 March 2011)

We report the observation of a gravitational-wave signal in data from a joint science run of the LIGO, Virgo and GEO 600 detectors. The signal exhibits the characteristic chirp waveform expected from a compact binary coalescence, and its form indicates a source with component masses $5.4 - 10.5\,M_\odot$ and $2.7 - 5.6\,M_\odot$ at a distance of less than 60 Mpc. There is strong evidence that the more massive component is a black hole with significant spin. The estimated false alarm rate for this event is 1 in 7 000 y, and detailed checks show no evidence that it is an instrumental artifact.

LIGO-P1000146-v16 – Circulation restricted to LSC and Virgo members

PACS numbers: 04.80.Nn, 04.25.dg, 95.85.Sz, 97.80.-d



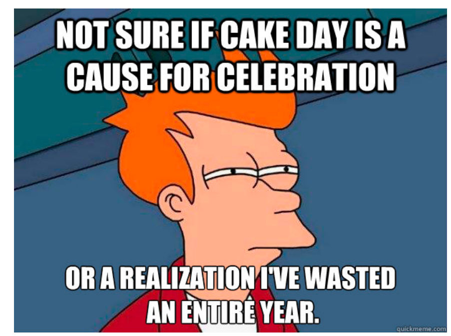
When Jay Marx opened his envelope in the LIGO Collab meeting in 2011

and told us all that the Big Dog was a Big Fake,

and that we had just completed the first successful discovery fire drill in gravitational wave observation history,

we still treated it as a moment of celebration.

We raised glasses of champagne, and toasted our fake success (insert appropriate emoji, if one exists).



The success: our collaboration had agreed for the first time what standards we'd use, and how we'd minimize our biases.

For the first time, we had decided what we need to have enough evidence for a detection and a great discovery.

Documenting the emotional stress

LIGO Scientific Collaboration

Detections | News | About | LIGO science | Educational resources

Latest news | News archive | Press releases | Press information | Blog | Awards & Recogn

"BLIND INJECTION" STRESS-TESTS LIGO AND VIRGO'S SEARCH FOR GRAVITATIONAL WAVES

MATTER | PHYSICS

The Astrophysicists Who Faked It

The inside story of the gravitational wave signal injection.

BY JONAH KANNER & ALAN WEINSTEIN
ILLUSTRATION BY BRIAN REA
NOVEMBER 3, 2016

http://nautil.us/issue/42/fakes/the-cosmologists-who-faked-it

Was GW150914 a blind injection?

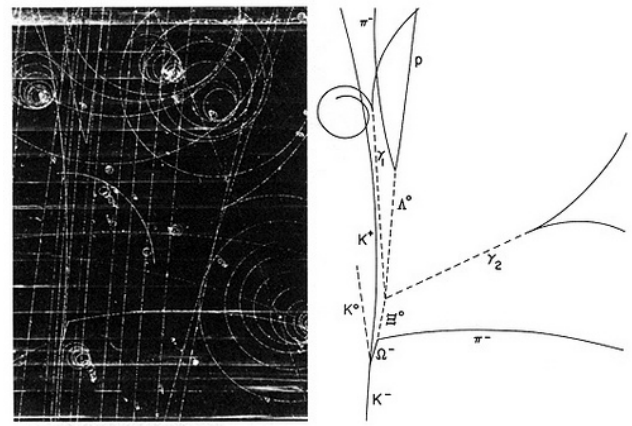
Fast forward to September 14, 2015:

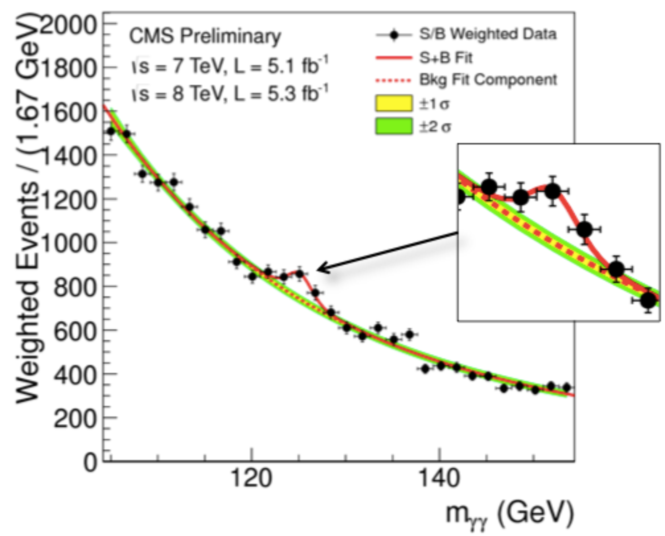
- We remembered the Big Dog saga, September 16, 2010. But …
- We were told: “there was no blind injection team convened for Advanced LIGO, and this wasn’t a blind injection.”
- In our electronic logbook: “There were NO Transient Injections during G184098 Candidate Event.” (award for best e-log entry of the year.)
- Still, we doubted this. We asked to look at the blind injection data channel. The response: “Go right ahead.” We looked and found nothing.
- Was GW150914 a blind injection?
  - » Check the not-so-secret blind injection channel. Nothing.
  - » A double-blind injection? Check the *secret* blind injection channel.
    - – There was no secret blind injection channel.
  - » A rogue injection, by a disaffected ex-employee? How to hide the evidence?
    - – We couldn’t figure out how to do it ourselves!
  - » An evil genius?
    - – Smarter than us!?

Blind analyses in LIGO GW searches

Discovery based on one event…?

- Rare in discovery physics
- cf discovery of the Ω^- at BNL in 1964,
- or 25 neutrinos from SN1987a
- Contrast with Higgs, 5σ peak in $m_{\gamma\gamma}$ above a large but smooth background





GW150914 was ONE event on top of an estimated background of $< 10^{-7}$ events (upper limit).

- We must minimize confirmation bias in establishing this number!

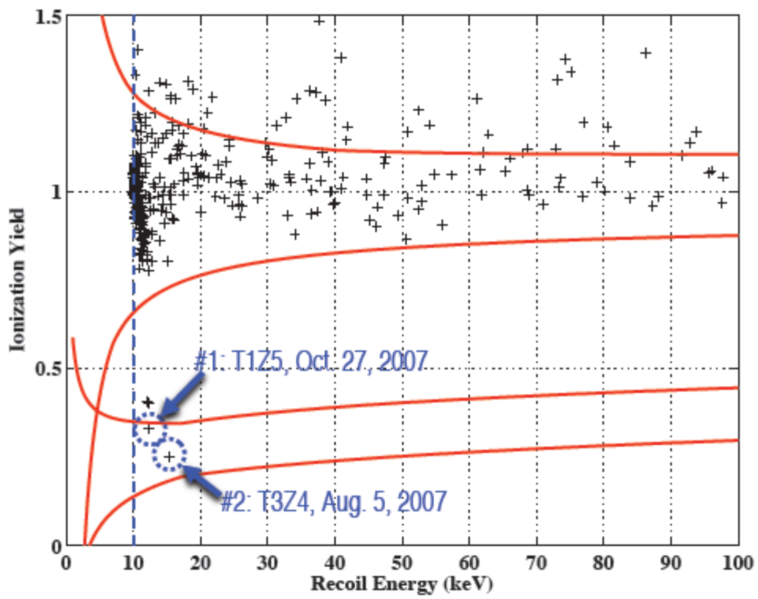
CDMS open box, 2007

WIMP-Search Data Set

3σ region masked
→ Hide unvetoed singles

Lift the mask, see 150 singles *failing* timing cut

Apply the timing cut, count the candidates: two events observed



Event #1 (T1Z5) shows no reconstruction issues

Event #2 (T3Z4) has a misreconstructed start time

A full reprocessing is needed to study this definitively

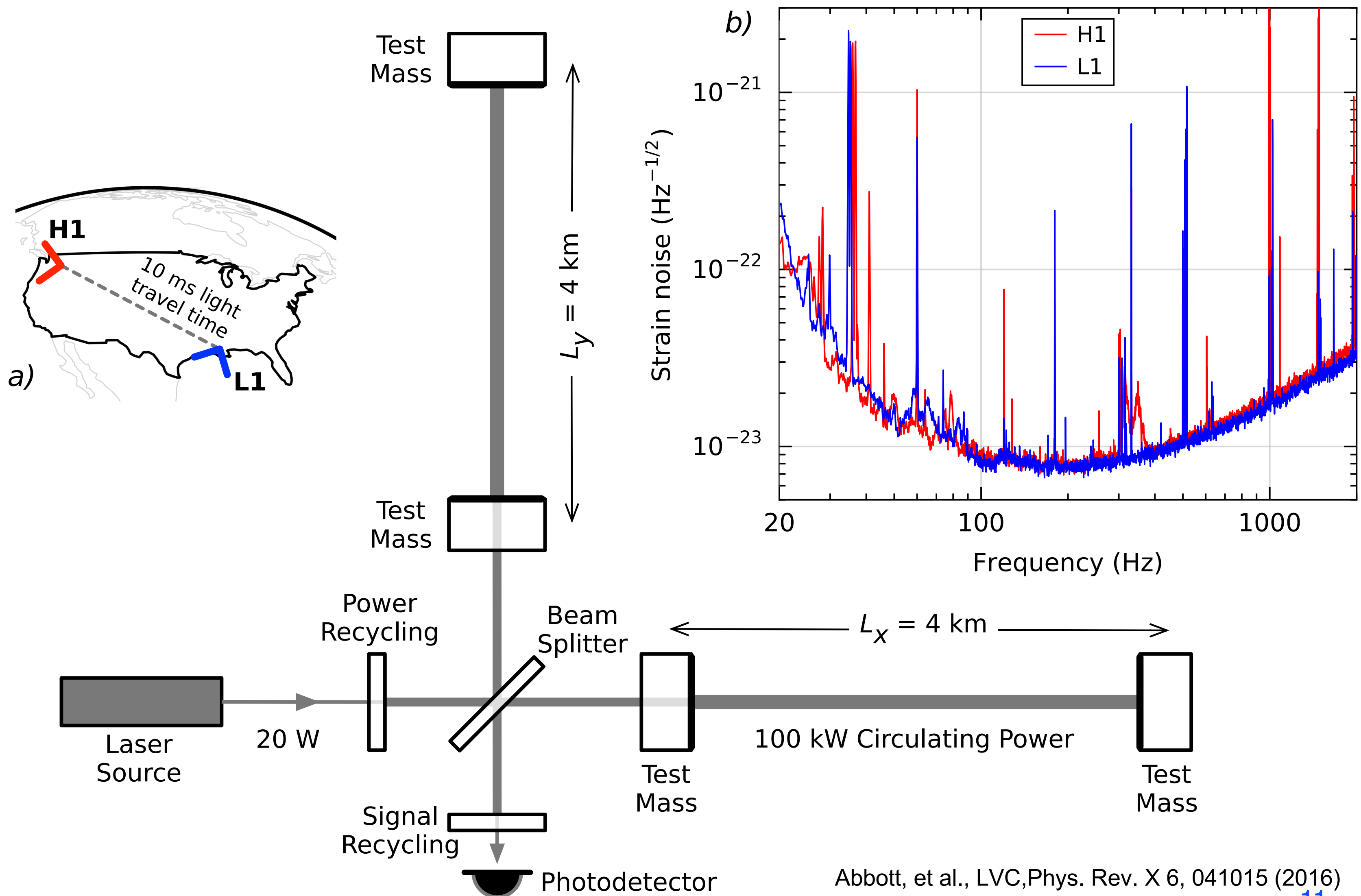
Background Estimate Redux

A refined estimate of the surface background accounting for this effect yields

Surface background
0.8 ± 0.1 (stat.) ± 0.2 (syst.)

With this revised estimate (and including neutron backgrounds), *the probability for observing at least 2 events is ~23%.*

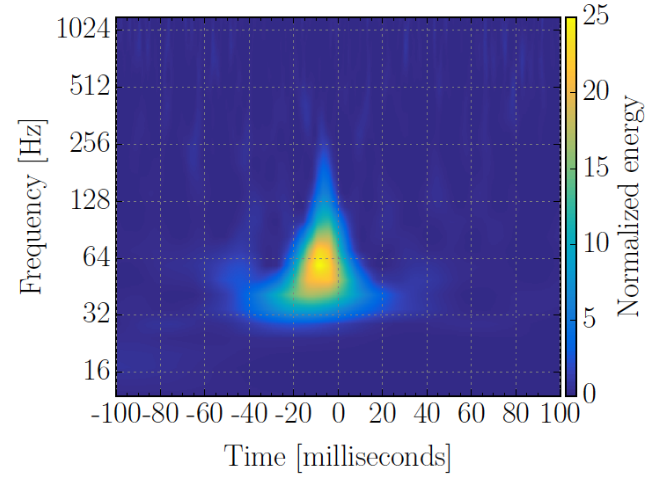
The Advanced LIGO detectors

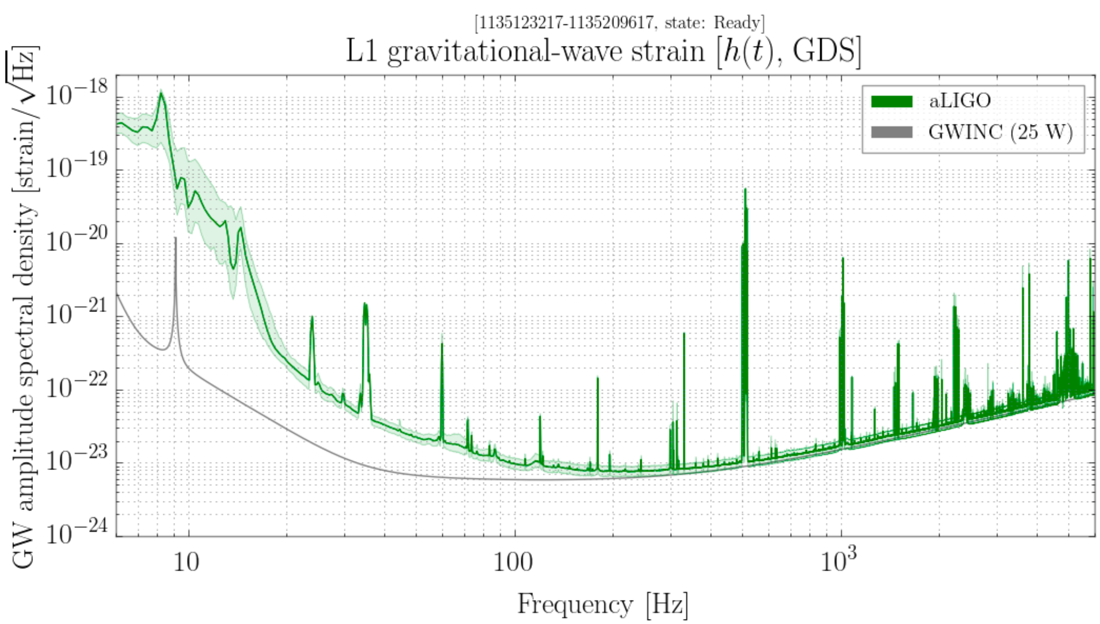


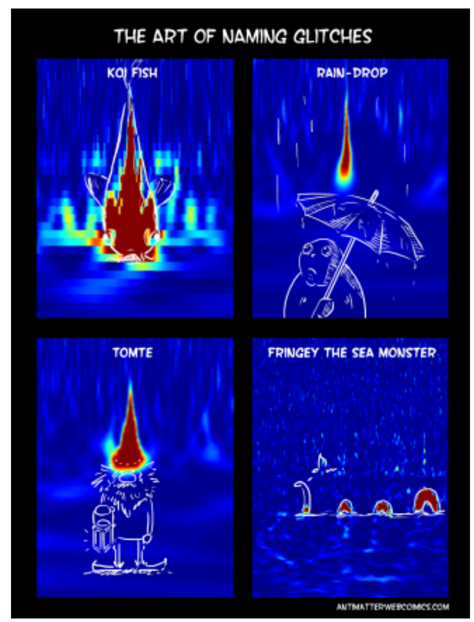
Abbott, et al., LVC,Phys. Rev. X 6, 041015 (2016)

The big issue: non-Gaussianity and non-stationarity of detector data

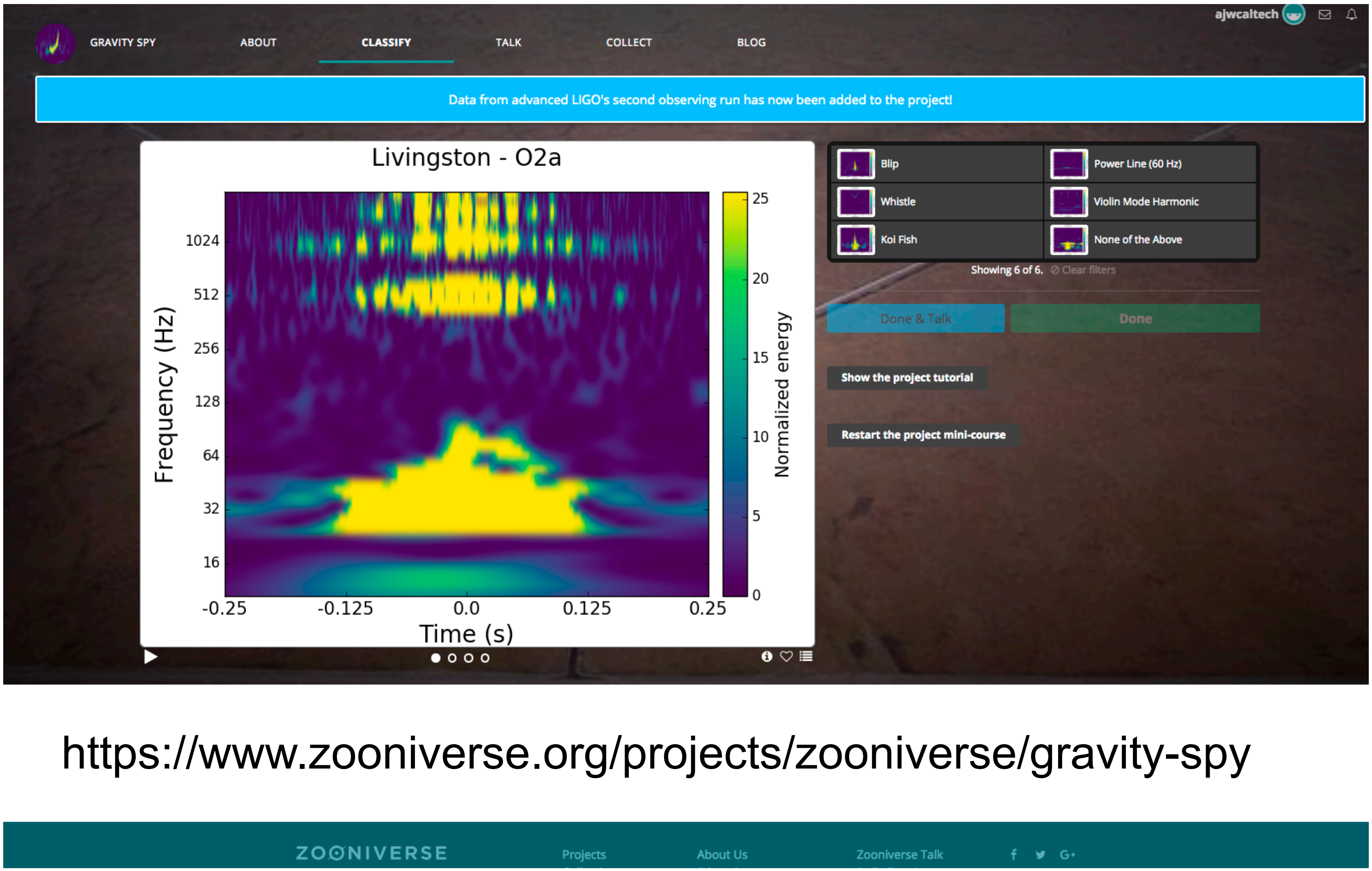
- Detector mis-behaviors could produce glitches that can be mistaken for an astrophysical signal (non-Gaussianity).
- Detector mis-behaviors change as we work to improve sensitivity … on ~ hour timescales! (non-stationarity)





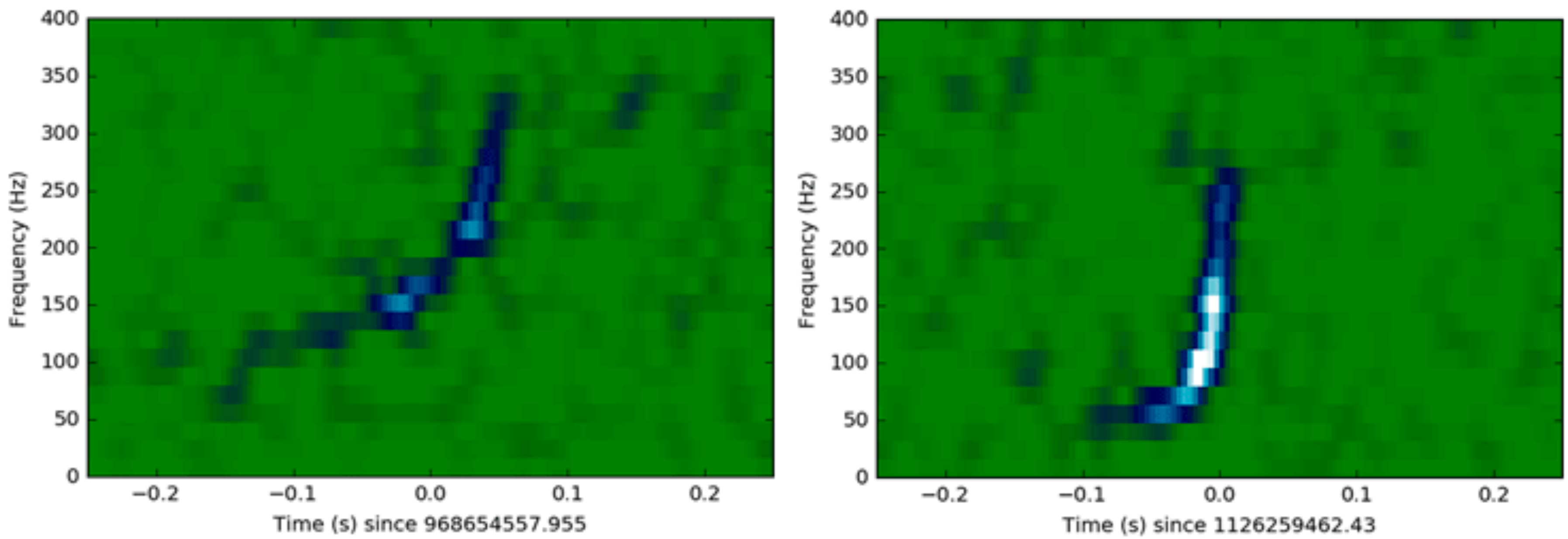


Classify LIGO glitches, and train a machine-learning algorithm!



https://www.zooniverse.org/projects/zooniverse/gravity-spy

Which is the real event?



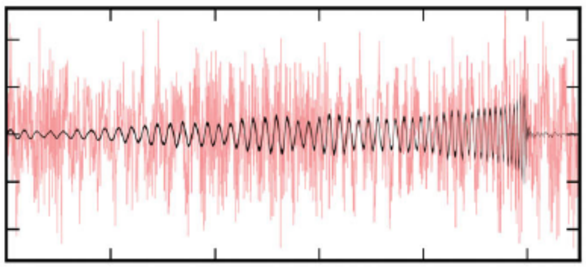
SPOT THE FAKE: The fake (left) and real (right) gravitational wave signals from 2010 and 2015, respectively.

Data from The LOSC Team and The LIGO Scientific Collaboration

It's very hard to imagine how detector glitches could fake a signal like this!
But … sometimes it's hard to tell ...

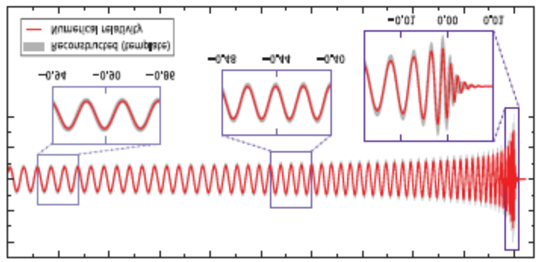
Matched Filtering

Data



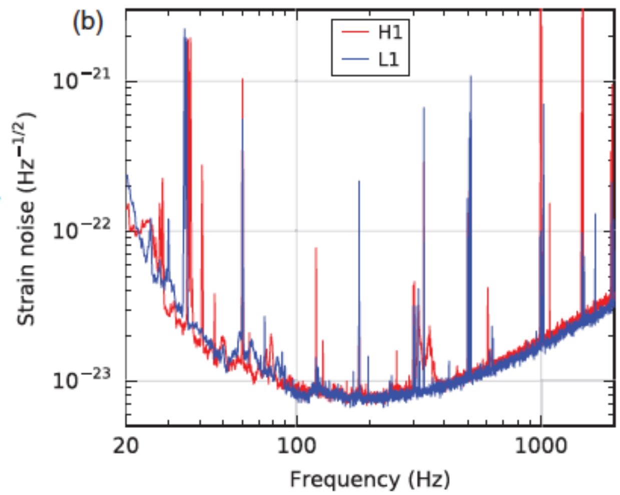
X

Waveform

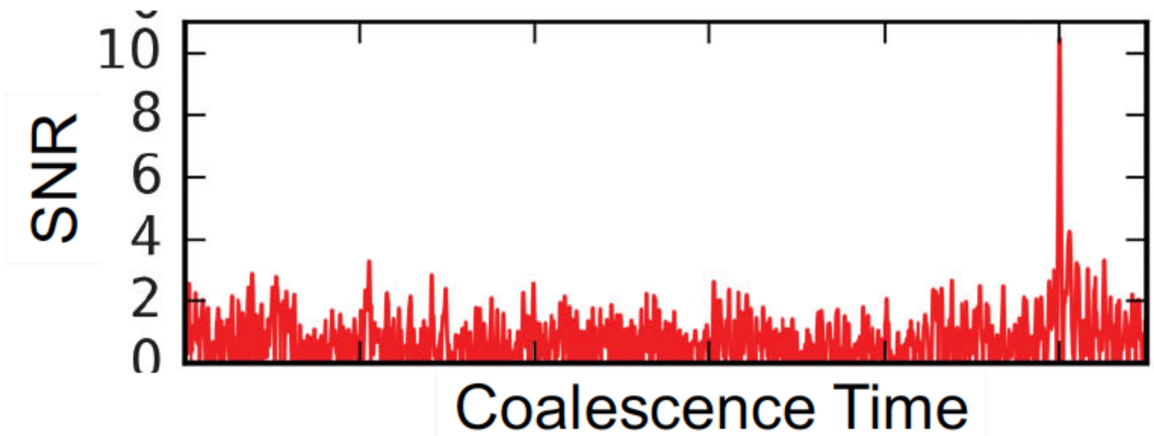


/

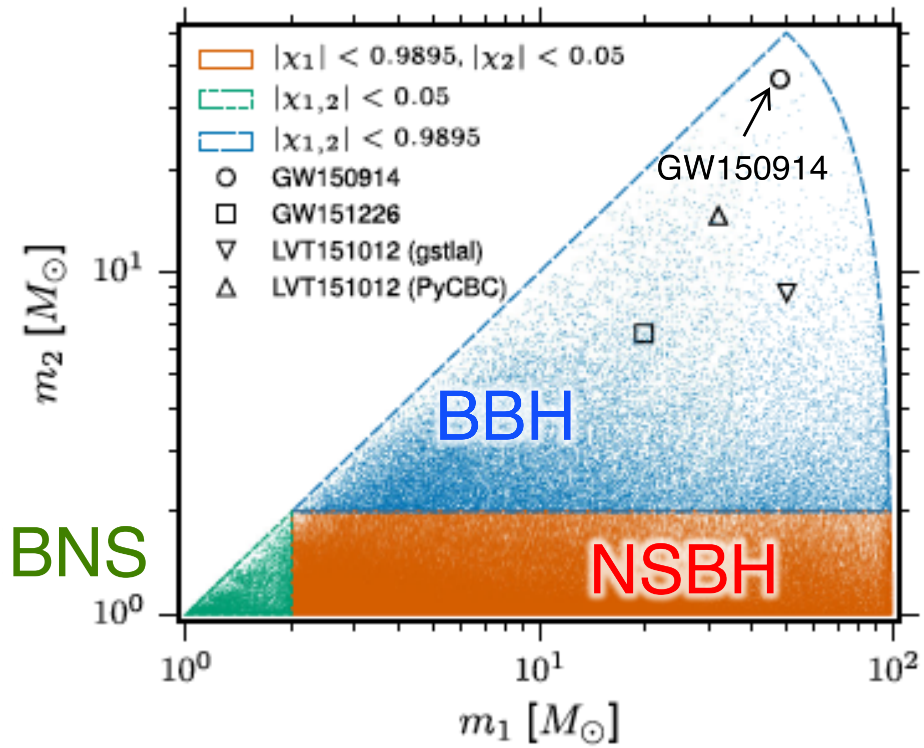
Sensitivity

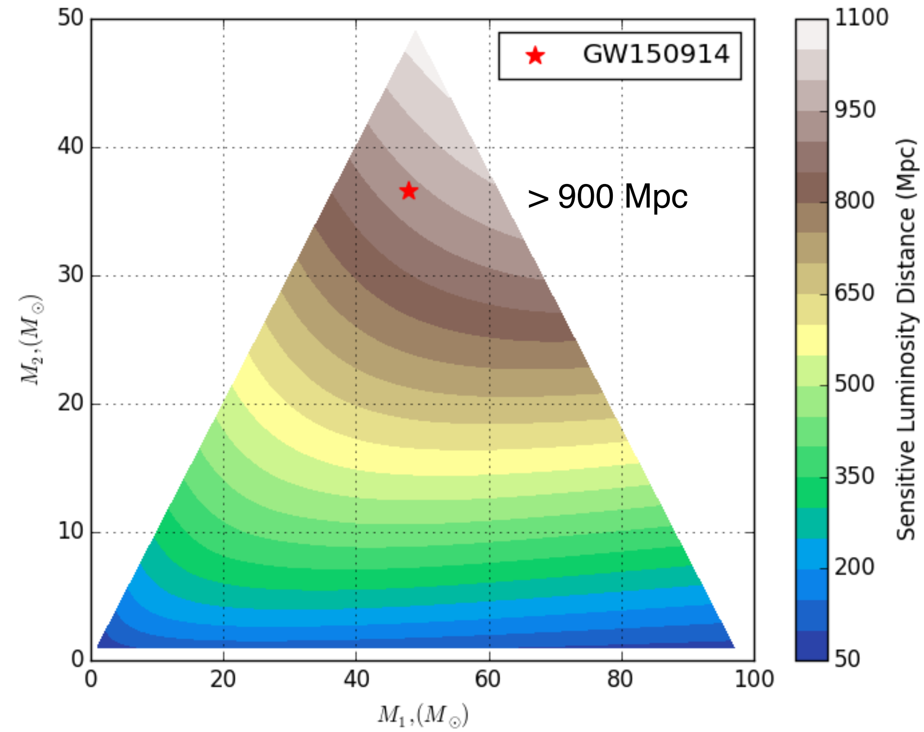


=



Template-based searches





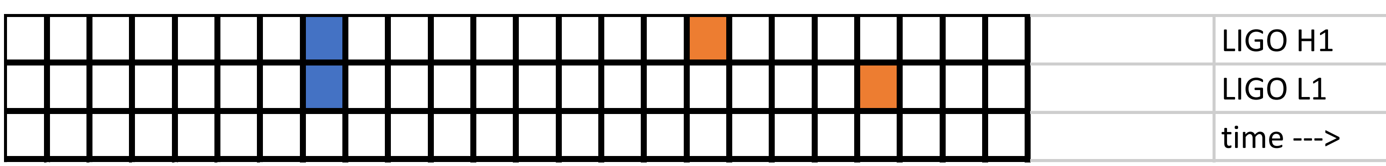
Masses and (aligned) spins
Templates spaced for < 3% loss of SNR: 250K templates.

Sensitive distance in Mpc

Abbott, et al., LIGO Scientific Collaboration and Virgo Collaboration, “Binary Black Hole Mergers in the first Advanced LIGO Observing Run”, https://arxiv.org/abs/1606.04856, accepted in Phys. Rev. X

Ranking the detection candidates

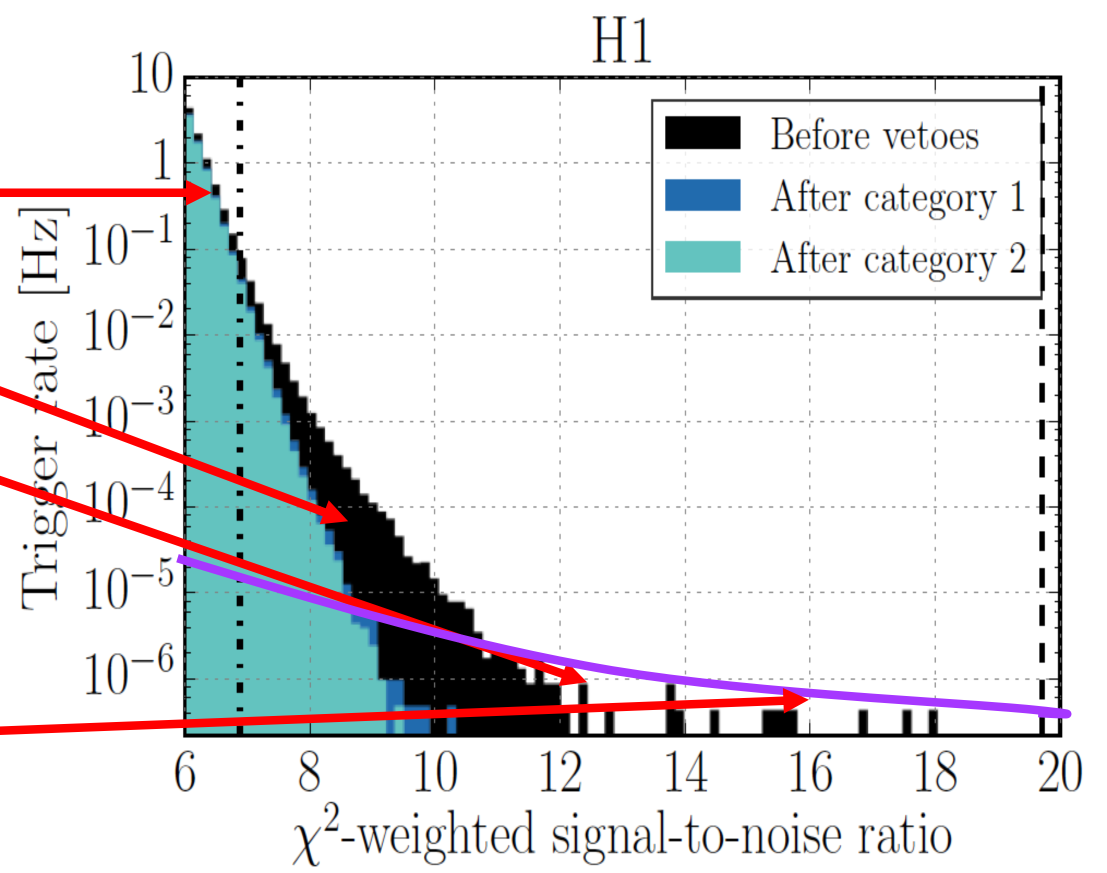
- Detector noise is of limited “loudness” (SNR)
- Rank events by loudness
- claim detection if an event is louder than the noise
- The challenge is to estimate the background under a loud event, if it is << 1 event
- Our most powerful tool: multi-detector coincidence
  - In the first Advanced LIGO observing run (O1, Sep 2015 – Jan 2016) we had two detectors: H1 and L1
  - Require coincidence in time (within light-travel time ($\pm$10 ms) and waveform model timing error $\pm$2 ms)
- Background estimated from time slides (“anti-coincidence”), which can’t come from real GWs



Candidate event trigger SNR distribution

- SNR $\rho \sim 1/\text{distance}$
- Gaussian core: $dN/d\rho \sim exp(-\rho^2/2)$
- Non-Gaussian tail
- Long tail, vetoed

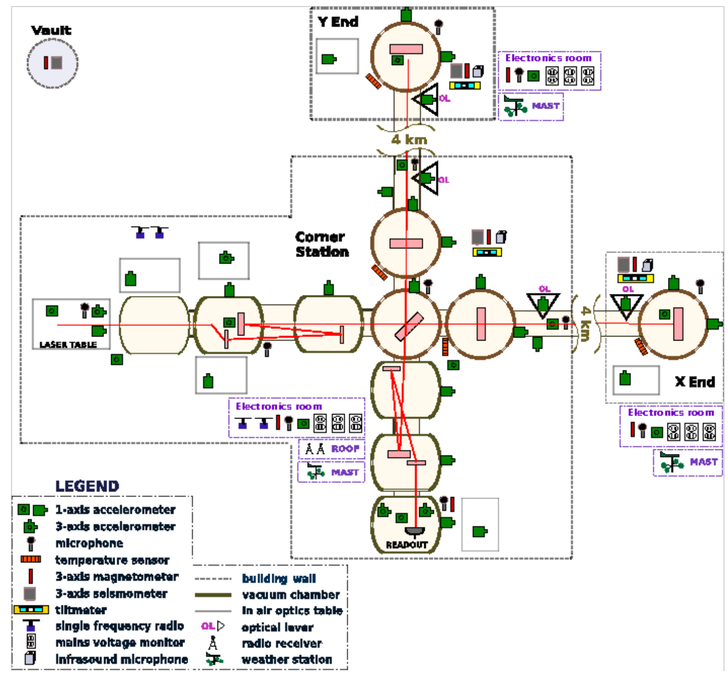
- The tail does not go on forever!
- Signal rate from cosmic events with uniform density in sensitive volume: $dN/d\rho \sim C / \rho^4$
- For high enough SNR, signal must dominate over noise!



(a)

Removing detector misbehavior "blindly"

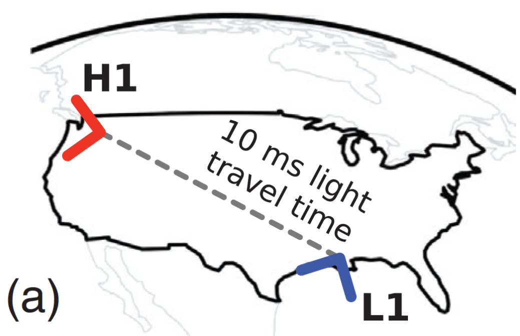
- Must veto non-Gaussian instrumental glitches, without bias
- How do we know that a glitch isn't a GW signal?
- Are most powerful tool – coincidence.
- Assumed uncorrelated between detectors – "anti-coincidence" time-slides can't be GW signals.

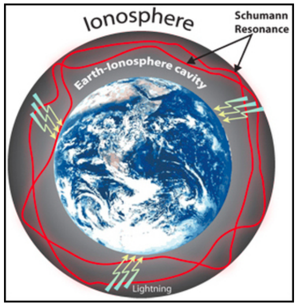


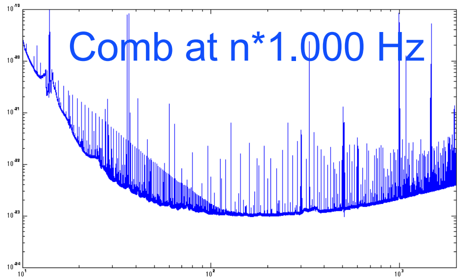
- We use these time-slide triggers to identify glitches in the strain data,
- and then look for correlations with data from our hundreds of physical environment monitoring sensors (seismometers, accelerometers, microphones, RFI monitors, voltage monitors, …)
- and thousands of interferometric detector monitoring data channels (everything but the strain data channel).
- From all of this, we define and veto times with clear (witnessed), or not so clear, detector misbehavior,
- *without ever looking at the coincident event triggers (closed box).*

Removing detector misbehavior "blindly"

- Assumed uncorrelated between detectors – "anti-coincidence" time-slides can't be GW signals.
- But … correlated glitches between detectors 3000 km apart:
  - Lightening storms – we monitor them, they don't produce glitches in the detector, and they don't look like signals in EM/RF detectors on site
  - Schumann resonances (global geomagnetic field)
  - GPS-based timing system – these can produce instrumental spectral lines!
- We have found *no evidence* for correlated transient noise between H1 and L1, and are convinced that this could not be responsible for our detected events.



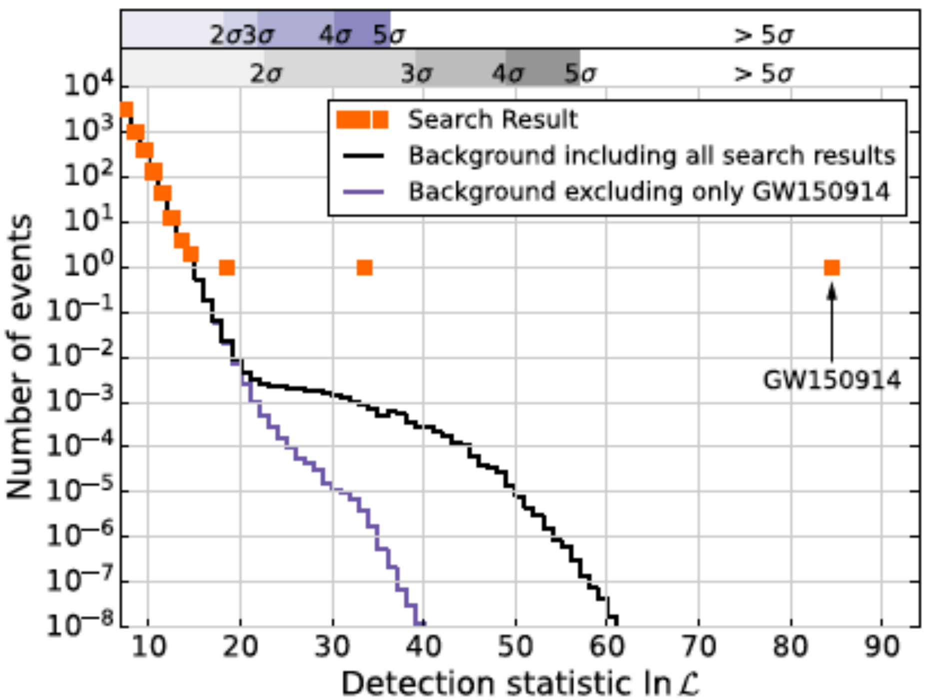


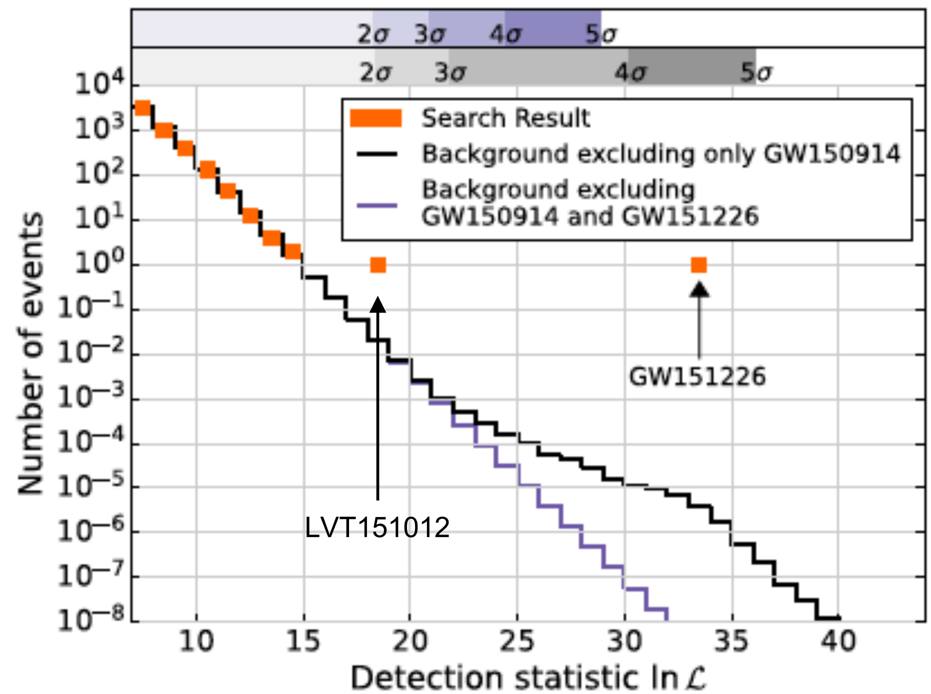


p-value vs. p-astro

- Our job is to convince ourselves, and you, that (free of bias) the event we see is not consistent with background from environmental or instrumental misbehavior.
- If the probability that an event is consistent with background (p-value, or False Alarm Probability FAP) is sufficiently small, you might conclude that it must be a real GW signal.
- But… what if GR is wrong, and GWs don't exist (and we've been fooled by the Hulse-Taylor binary)?
- Then, no matter how small the p-value, the event *is* background, with certainty!
- We need to quantify the *prior* on the existence, and rate, of GWs from the sources that we are considering.
- Until we can estimate the rate (ideally, through direct observation), we don't know the probability that the event is astrophysical.
- We've got a chicken-and-egg problem.
- Fortunately, that's only true for the *first event*. One event gives us an estimate on the prior, which we can use to compute p-astro for the 2nd event, and so on.

Opening the box: Search results Advanced LIGO Observing Run O1





- Tune analysis pipeline and data quality vetoes "blindly", using time-slid data to estimate the background, and software injections for the signal.
- All tuning finalized before "opening the box" on in-time coincident triggers.
- **We find three events above estimated "background" from accidental coincidence of noise triggers.**
- Two have high significance (> 5σ).

Abbott, et al., LIGO Scientific Collaboration and Virgo Collaboration, "Binary Black Hole Mergers in the first Advanced LIGO Observing Run", https://arxiv.org/abs/1606.04856, accepted in Phys. Rev. X

The "little dogs"

- Background estimation makes use of data that include signals!
- Time-slides don't eliminate this possibility
- We assume that the signal population is much smaller than the noise trigger population; single-detector triggers are overwhelmingly noise
- But for a loud event like GW150914, the loudest single-detector triggers, and the loudest time-slide coincident triggers, are *associated with the event*.
- And we know which ones they are! So we can remove them.
- But should we? Removing them is an un-blinding bias, and a particularly dangerous one.
- But leaving them in is crazy, no?

To kill the “little dogs”, or not?

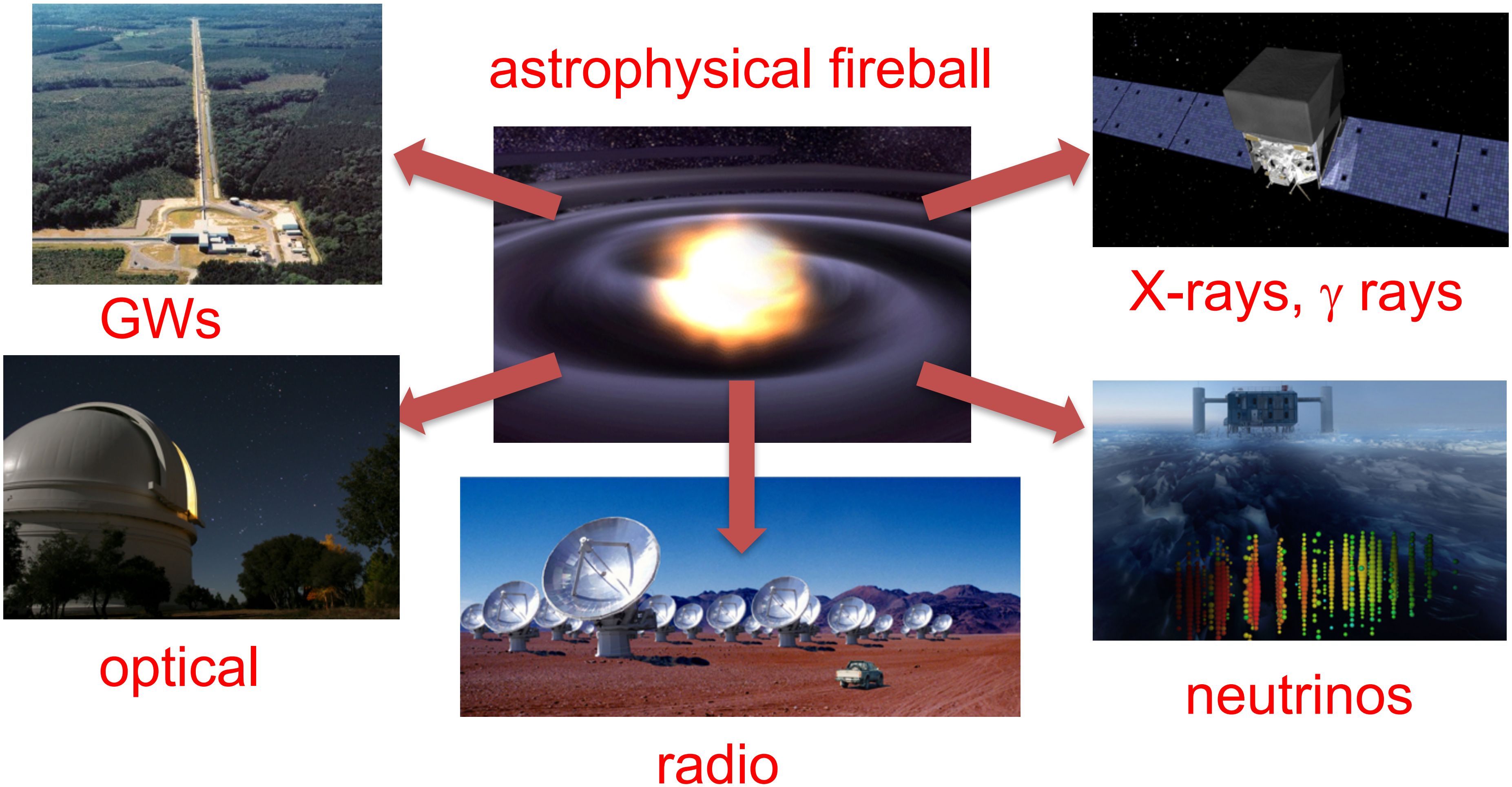
- But leaving them in is crazy, no?
- There is no right answer to this question (that I know of).
- All we can do is “bracket” the true best background estimate: leave in the little dogs, or take them out.
- Can be repeated hierarchically
- First observation: Is it signal or is it background?
- **This is a deep problem in discovery physics!**

First observation is a discovery! 😍
Second observation is a confirmation 😛.
Third observation is a distribution 🙄.
Fourth observation is a calibration😒.
Fifth observation is background 😠

Accumulating sufficient statistics for background estimation

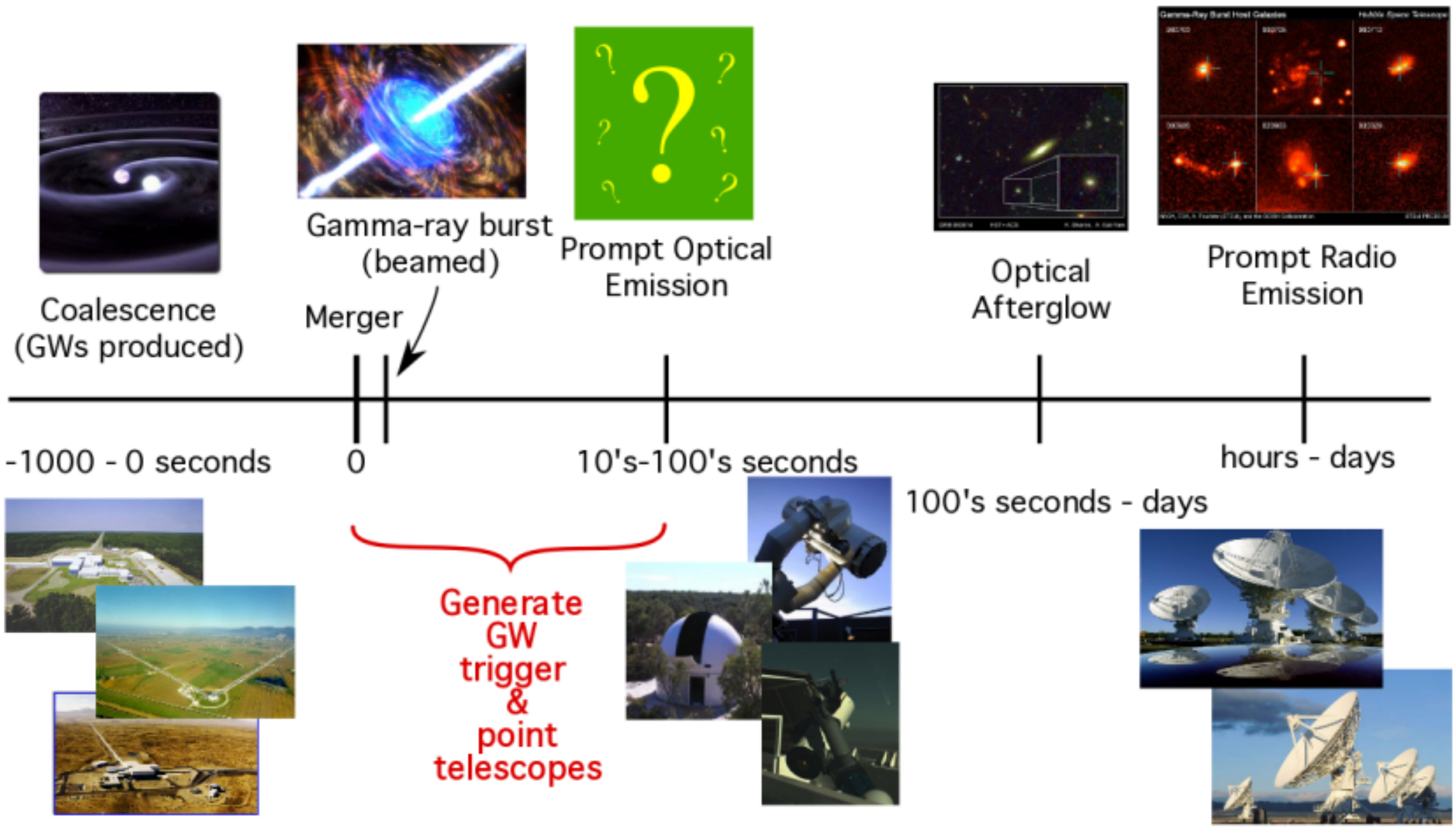
- To get enough statistics in time-slides to call *one* event a 5σ detection, we need ~ 10^7 trials for each observation.
- We require time-coincidence within ~ 0.01 s, but a 0.01 s time-slide is too short, because our waveforms have significant auto-correlation out to 0.1 s. We dare not time-slide any shorter than that.
- In 10 days of double-coincident observing, we have roughly 10^7 un-correlated double-coincidence trials, and $(10^7)^2$ time slides.
- We dare not go longer, because the detector noise is *non-stationary* and can change noticeably (especially for the loudest glitches) over ~10 days.
- So… we have to wait for 10 days to collect enough noise triggers to estimate the background for a loud event to claim 5σ detection, (FAP < 10^{-7}, FAR < 1/200,000 yrs), *before opening the box*.
- Can we wait that long? Well, yes and NO!

Multi-messenger Astronomy with Gravitational Waves



Low-latency identification of transients for rapid (< ~100s) followup

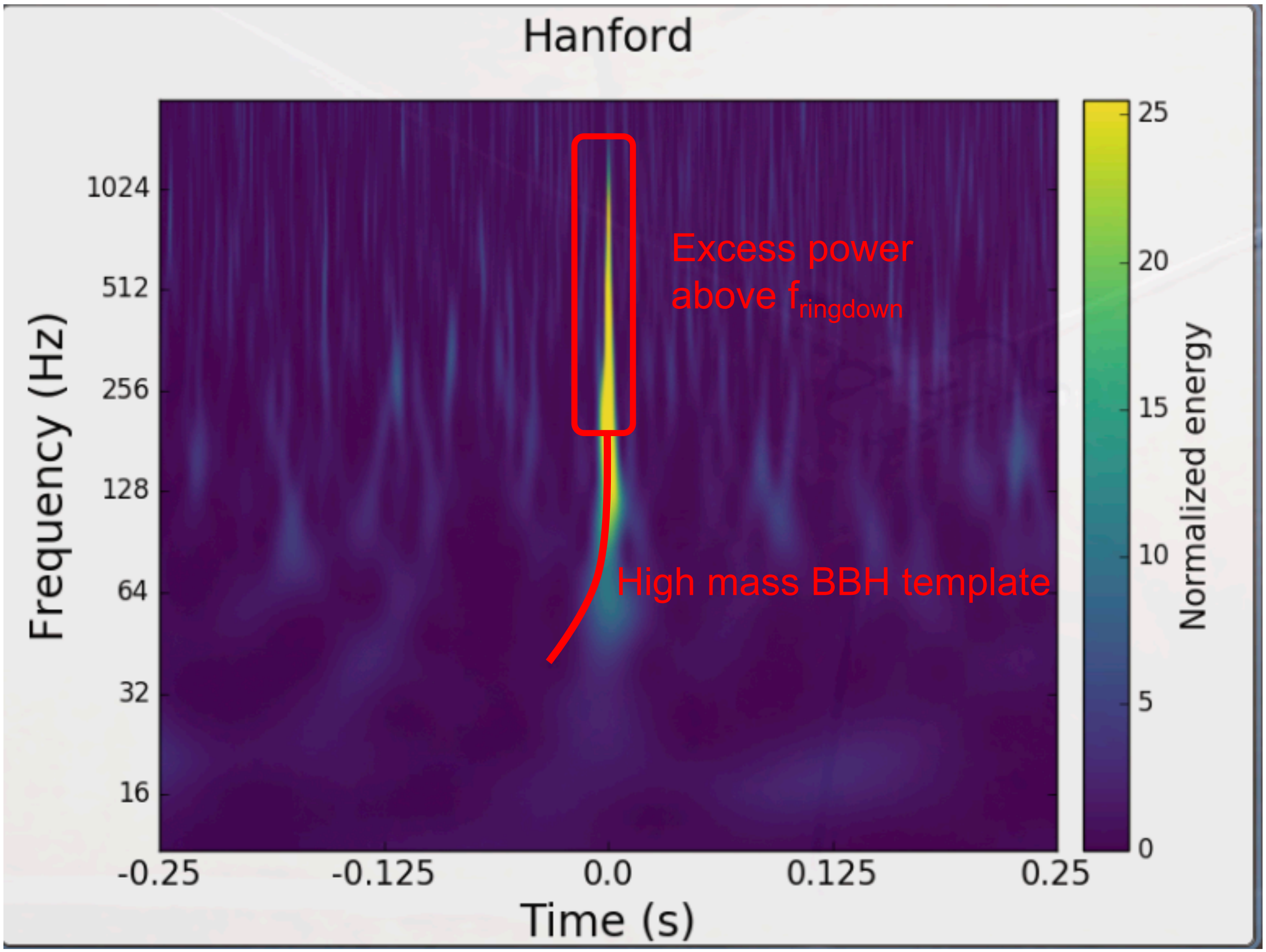
EM counterparts to GW sources (if any) are short-lived and faint



Are we really doing a blind analysis?

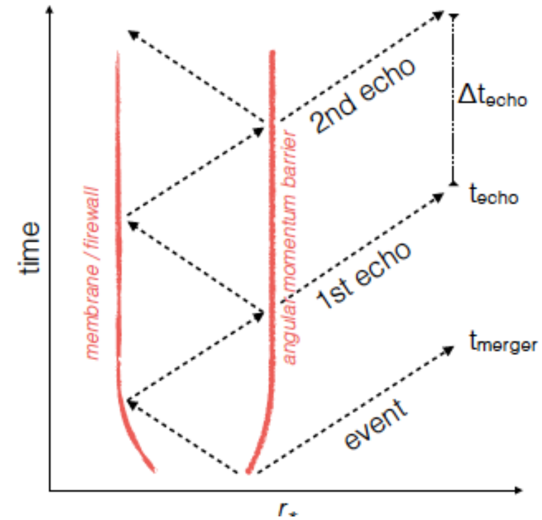
- Can we "open the box" at low latency (to alert our EM partners), then close it again for 10 days, then open it again, and claim that we are doing a proper blind analysis?
- Can we automate the data quality procedures to make them "blind" to anything we see at low latency?
- Well, we can try. But data quality investigations often lead to new kinds of detector misbehavior that must be dealt with before opening the box again.
- In the end, a purist would conclude that *we are not really* conducting a true blind analysis.
- In our case, discovery science means not only discovering new signals, but also *discovering new backgrounds* (detector misbehavior).
- In some sense, our (engineered) detectors may be more complicated to understand than binary black hole mergers 🤔
- They are the "evil genius" that could fool us!

How to get it wrong: high mass binary black holes, higher order modes



How to get it wrong: echoes from the abyss

- When is a BH *not* a BH?
- Planck-scale departures from GR (firewalls, fuzzballs, gravastars) near the putative BH horizons can lead to “echoes”.
- repeating damped echoes with time-delays of *8M logM*
- Abedi, Dykaar, and Afshordi, arXiv:1612.00266v1
- “… we find tentative evidence for Planck-scale structure near black hole horizons at 2.9σ significance level”
- But… if you look for ringdowns in LIGO strain noise, you will find it *everywhere*
- (They used 32s of data around GW150914 to estimate the background).



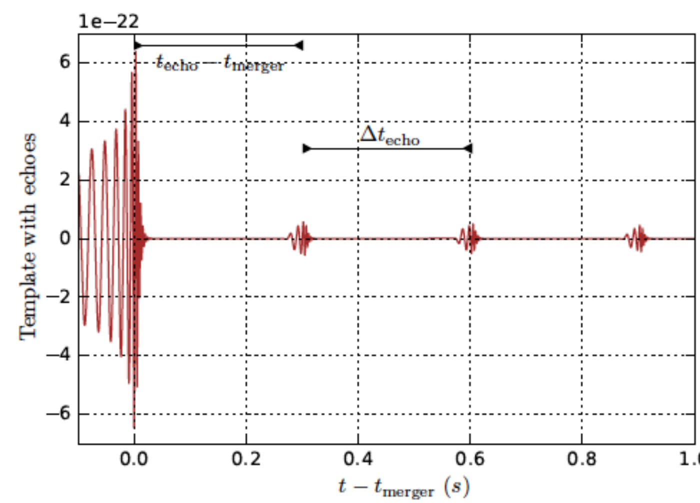
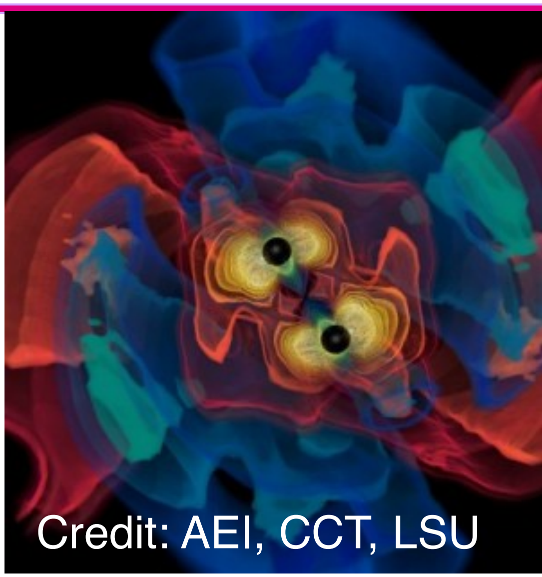


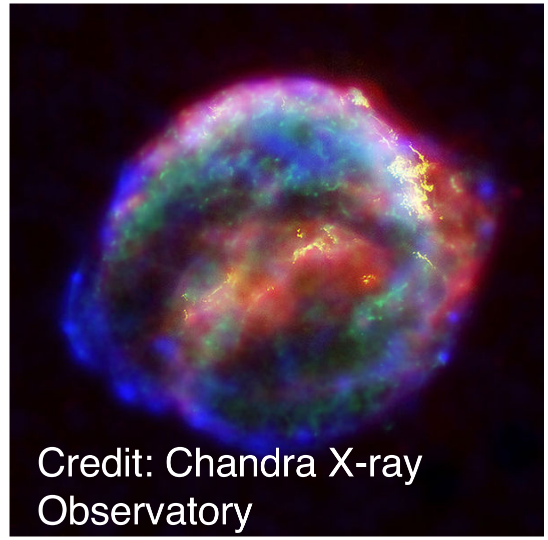
FIG. 2: LIGO original template for GW150914, along with our best fit template for the echoes.

GW sources for LIGO: The most energetic processes in the universe



Coalescing Compact Binary Systems: *Neutron Star-NS, Black Hole-NS, BH-BH*

- Strong emitters, well-modeled,
- (effectively) transient

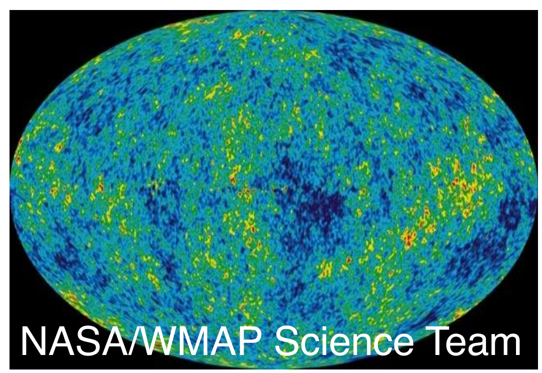


Asymmetric Core Collapse Supernovae

- Weak emitters, not well-modeled (‘bursts’), transient
- Cosmic strings, soft gamma repeaters, pulsar glitches also in ‘burst’ class

Spinning neutron stars

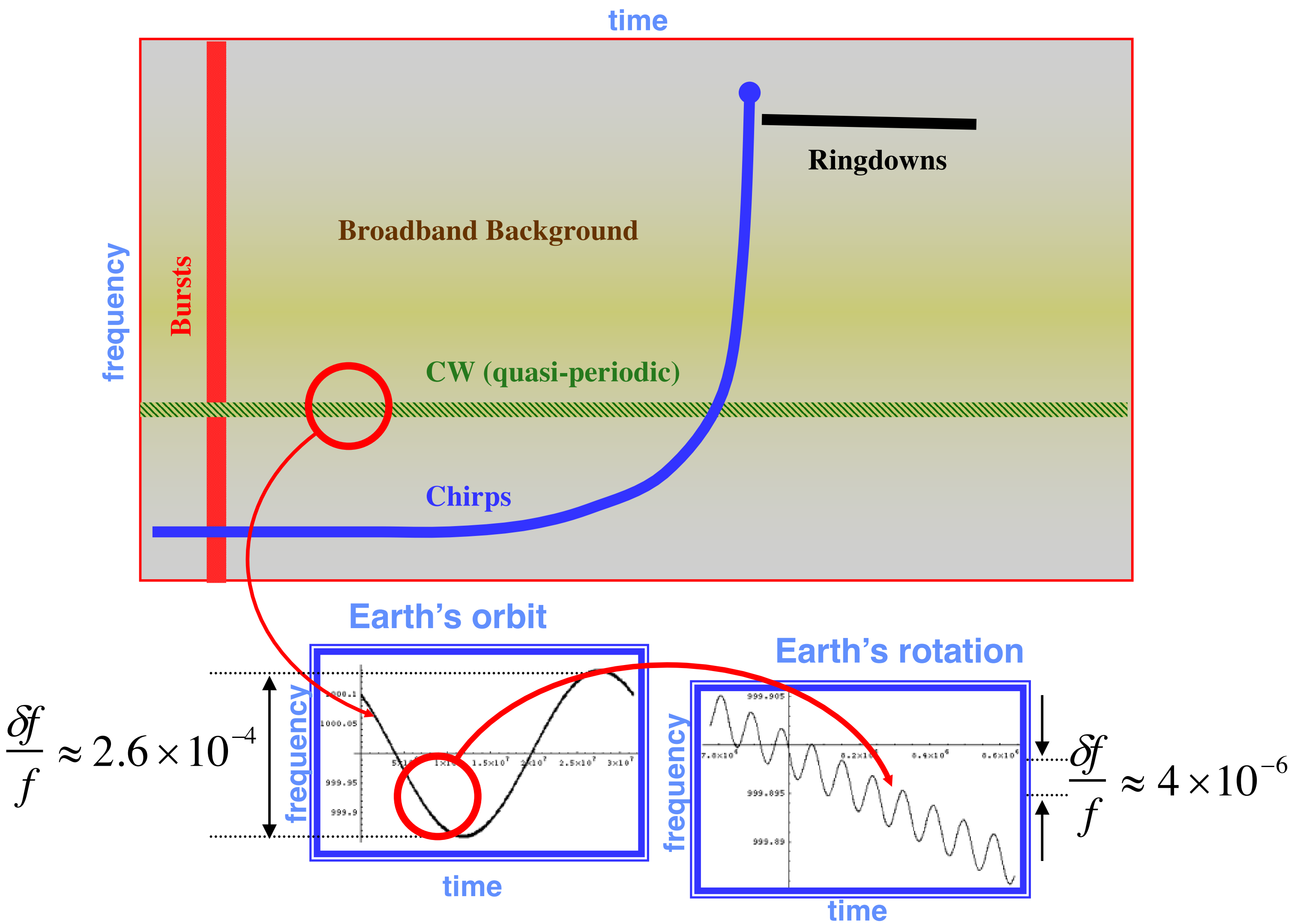
- (effectively) monotonic waveform
- Long duration



Cosmic Gravitational-wave Background

- Residue of the Big Bang, long duration
- Long duration, stochastic background

Frequency-Time Characteristics of GW Sources



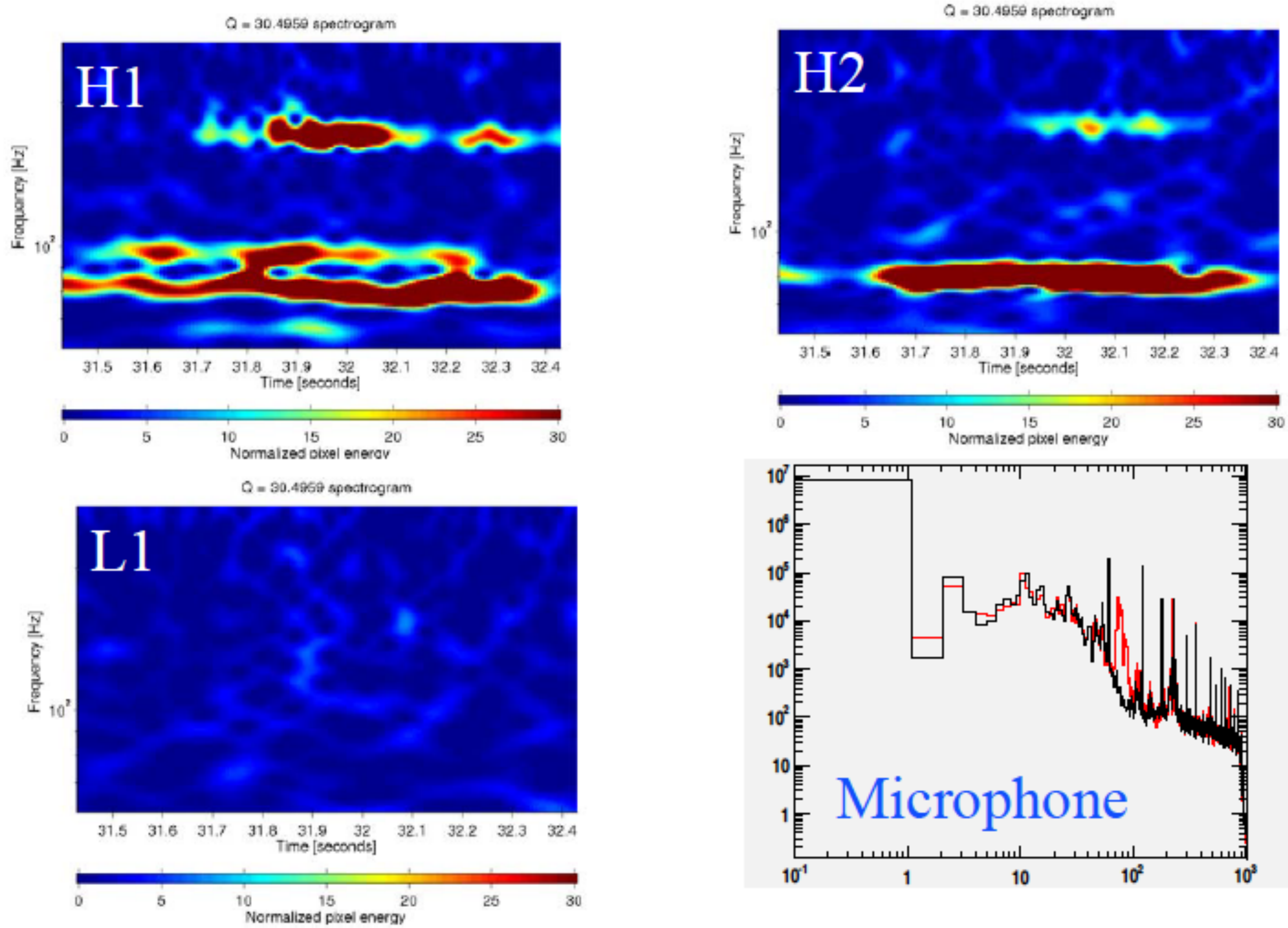
"Unmodeled" Transient GWs (Bursts)

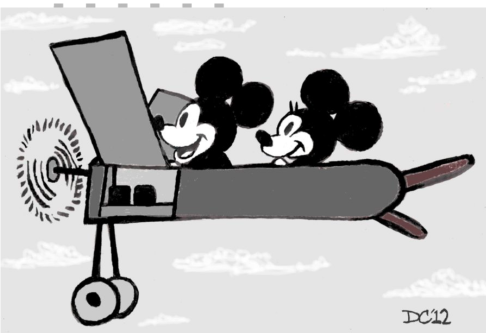
- From, eg, Galactic CCSNe, magnetar flares/starquakes, …
- Short duration, potentially broad-band; … unknown!
- A starquake might produce a narrow-band "ringdown" waveform. It is very easy to find ringdowns in aLIGO detector noise!
- Outliers from a "blind" search pipeline are scrutinized by hand for any evidence of instrumental misbehavior. Very hard to automate in a blind way.
- Coincidence with similar morphology is our most powerful tool.
- But… we cannot claim to be doing a truly blind search.

GW Burst candidate event from 2003

- Initial LIGO's second Science Run (S2) from Feb 14 - Apr 14, 2003
- We prepared a **rigorously blind** analysis pipeline / procedure for identifying GW burst candidates,
- with a pre-defined threshold on signal significance, based on excess power in the t-f plane,
- set to a false alarm probability (p-value) of 0.05.
- Much care went into data quality vetoes and methods to suppress detector noise glitches, understanding detector performance and misbehaviors from single-detector triggers and time slides.
- **One event** appeared in the open box, at triple coincidence (H1, H2, L1; with H1 and H2 co-located in the same vacuum envelope).

Loudest event in S2 burst search, triple coincidence





The Airplane Event

- It was rapidly identified as acoustic noise from a low-flying propeller plane over Hanford, somehow getting into the strain channels of both H1 and H2, accidentally coincident with a small upward fluctuation in noise in L1.
- Could hear the plane in microphones throughout the site;
  easy to veto post-facto.
- Why did we not have a veto in place already?
  Private planes are not allowed to fly over Hanford! The pilot was lost!
- Discussion of whether to exclude or include the candidate event in computation of upper limits spurred a **protracted “religious war”** between blind analysis statistical purists, and people who thought (post-factor) that it was too premature to apply strict blinding,
- including all the instrument scientists who invented, designed, built, optimized and operated the detectors, who thought the analysists were crazy people.
- Subsequent runs: more careful acoustic vetoes; optical sensing systems in acoustically-shielded “meat-lockers”. IN Advanced LIGO, they are in vacuum on seismic isolation platforms.

Continuous GWs from spinning neutron stars (pulsars)

- These are not events in time, they are "events" in frequency (spectral lines).
- They are expected to be very weak, visible only after long FFTs and averaging over months of data.
- They are Doppler modulated with sidereal and annual periods, which must be corrected for.
- LIGO detectors have lots of spectral lines from terrestrial (instrumental and environmental) sources.
- Many are strong, and don't have the unique sidereal modulation.
- Many are weak (combs), and/or do have solar-day and annual modulation.
- Outliers from a "blind" search are scrutinized by hand; we cannot claim to be doing a blind analysis.
- "When in doubt, take more data!"

Take-home message

- Discovery science is tricky!
- Doing an un-blinded search carries much opportunity for subtle biases – dangerous!
- It really helps to have a very well-defined model of the signal being searched for,
- and a really well-understood detector,
- and redundancy (coincidence between multiple detectors),
- and a really robust estimate of the background,
- and hope that you can *really* tell the difference between signal and background reliably and robustly, even before opening the box.
- Discovery science often doesn't come so simply …
- So it can happen that doing a truly blind analysis is dangerous!
- And … beware the evil genius!