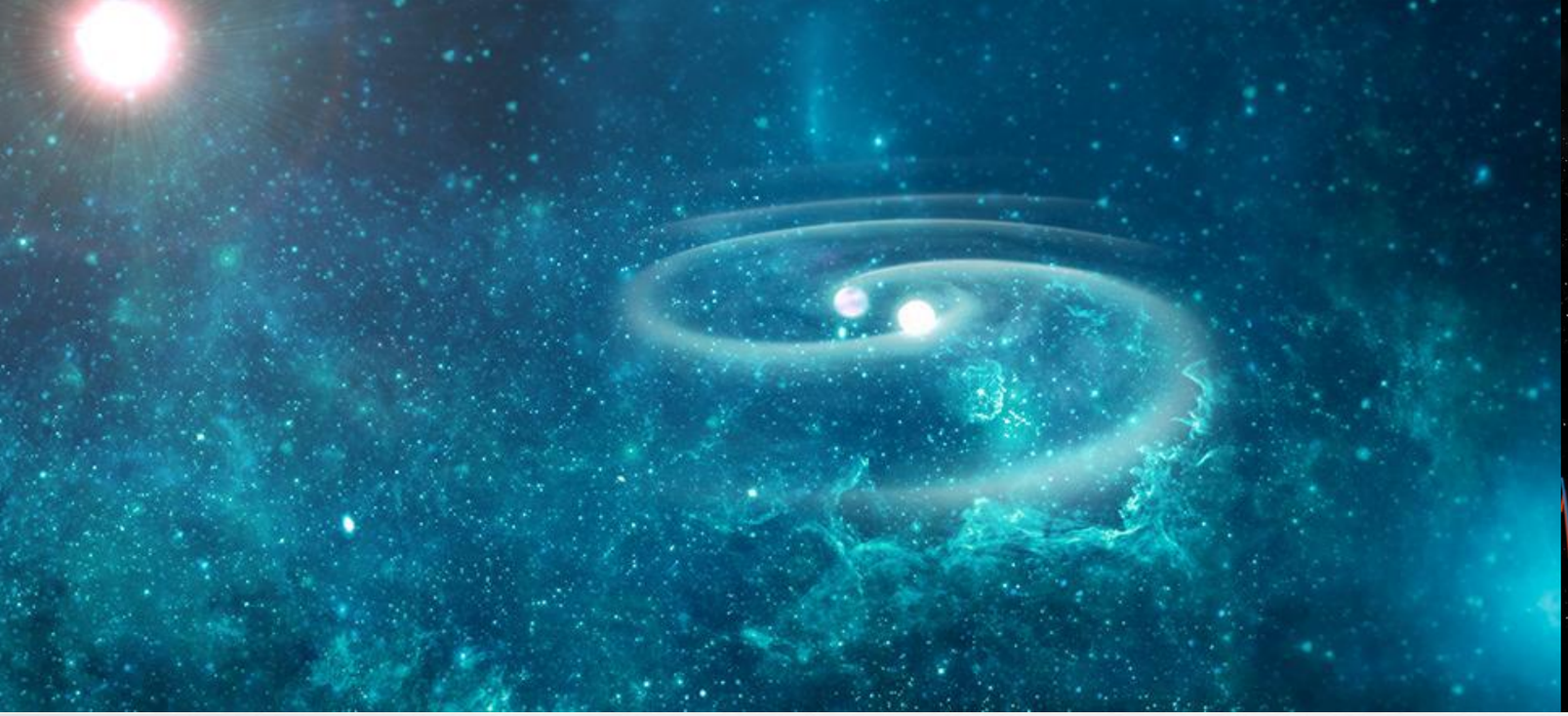
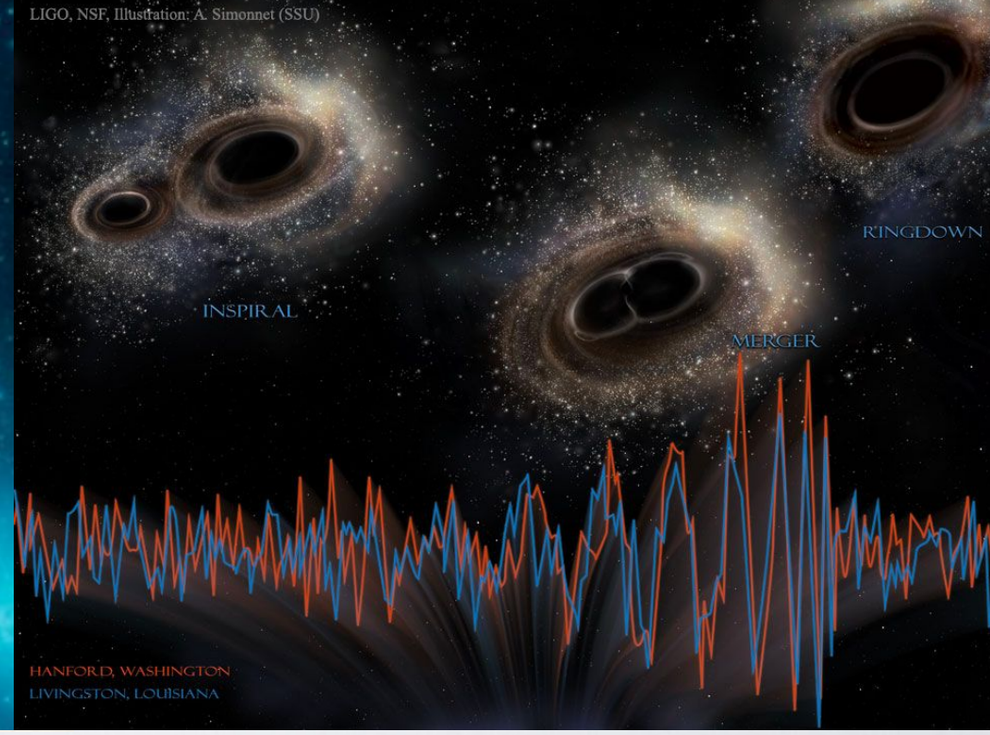




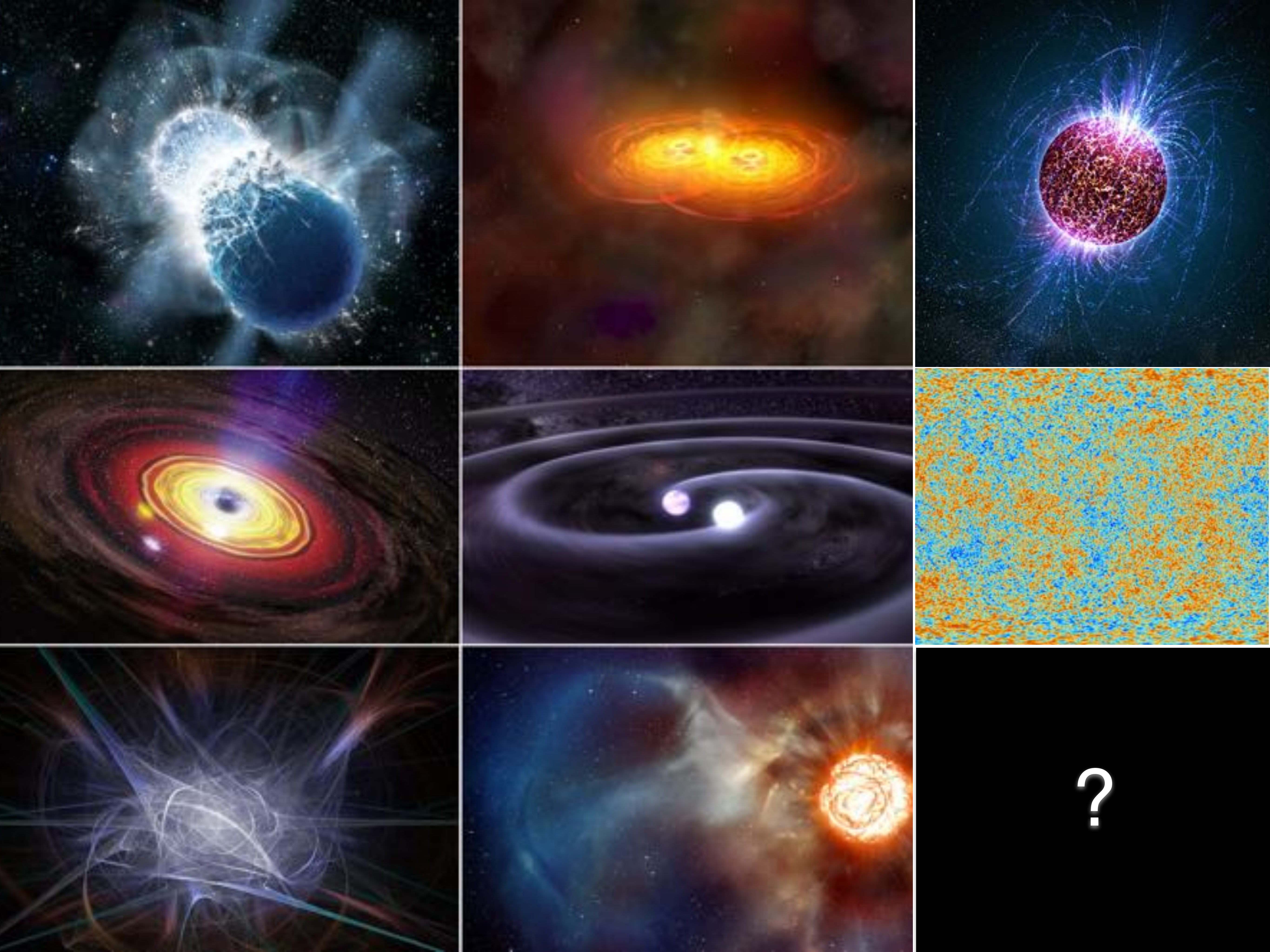
DATA ANALYSIS METHODOLOGIES

PhD International School on Technologies in Gravitational Wave Detection
Erice – May 25, 2026

Francesco Pannarale

SAPIENZA
UNIVERSITÀ DI ROMA

INFN
ROMA
Istituto Nazionale di Fisica Nucleare
Sezione di Roma



Gravitational-wave detectors

Galactic binaries

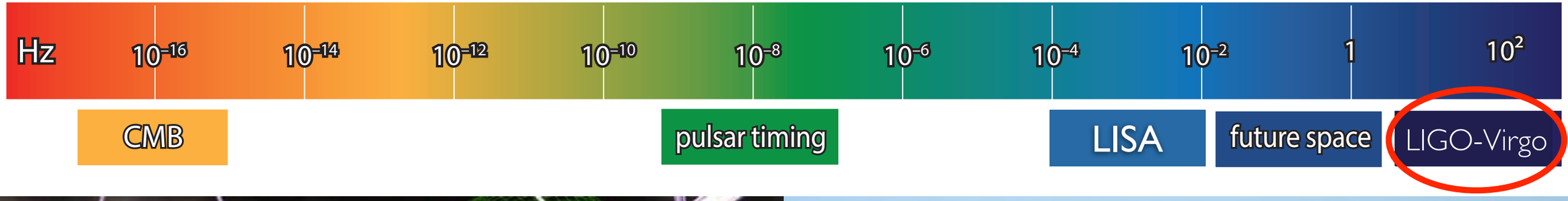
captures into MBHs

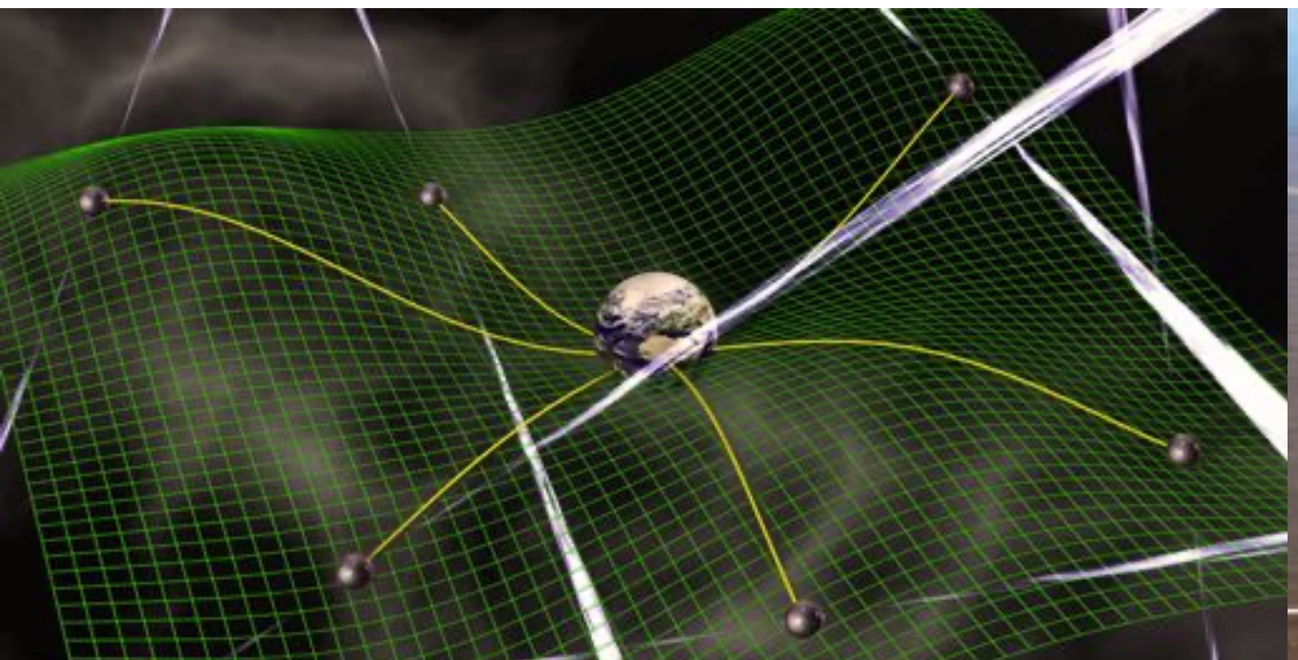
rotating NS

merging NS, BH

massive black-hole binaries

early-Universe quantum fluctuations



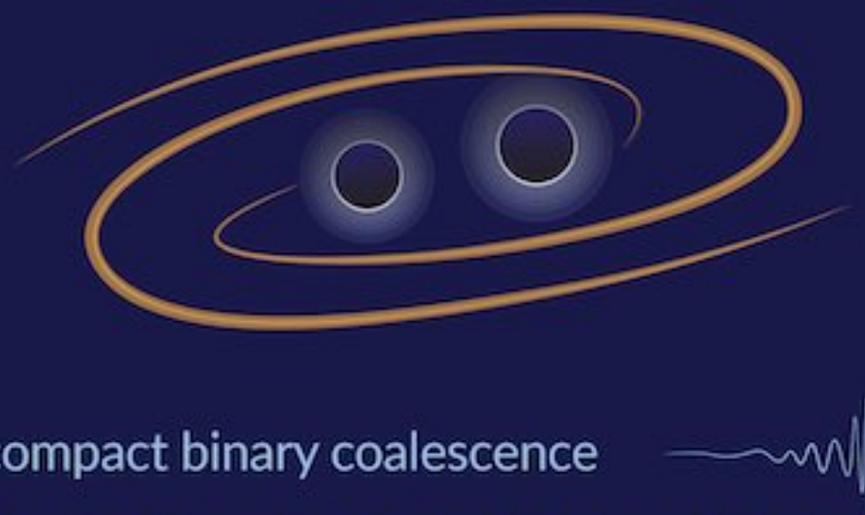


Four Data Analysis Areas

Modelled

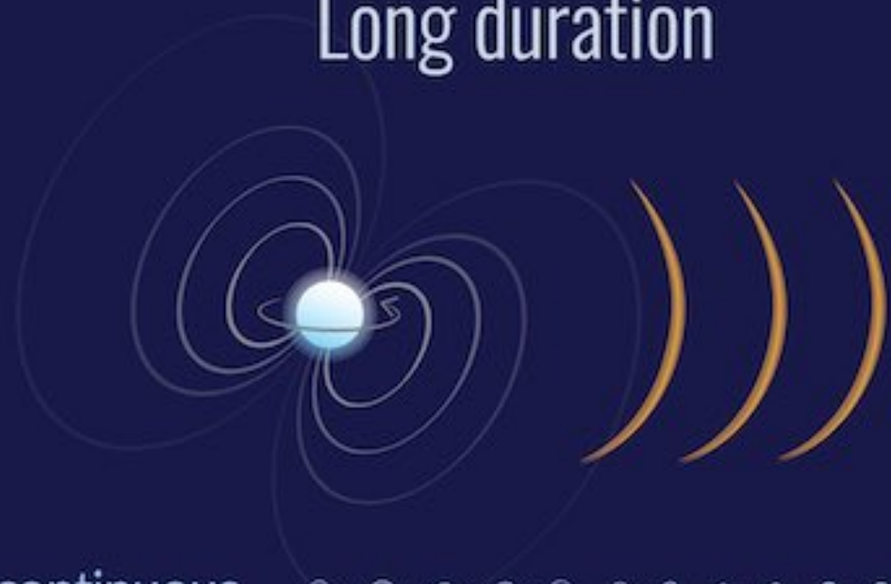
Unmodelled

Short duration



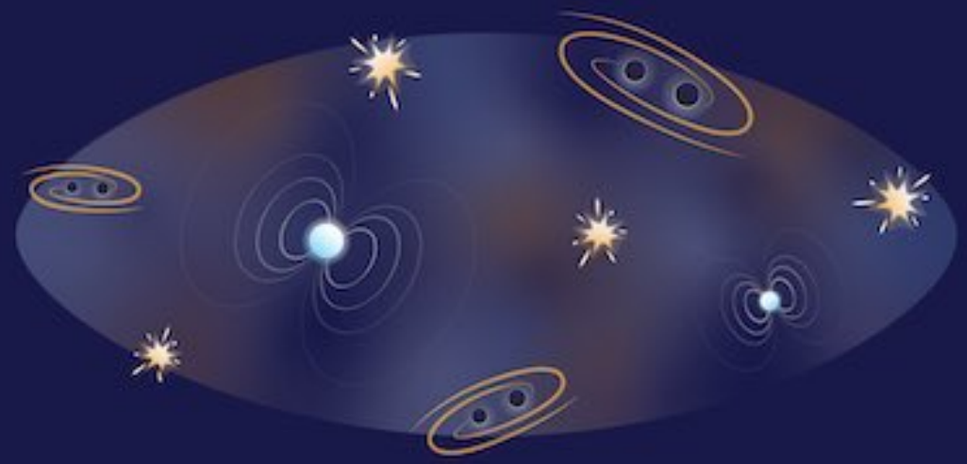
compact binary coalescence

Long duration



continuous

burst



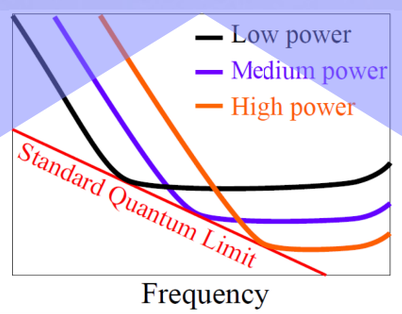
stochastic

GW Detector Data

- Time series of spacetime strain measurement, sampled at ~10 kHz, contaminated with noise:

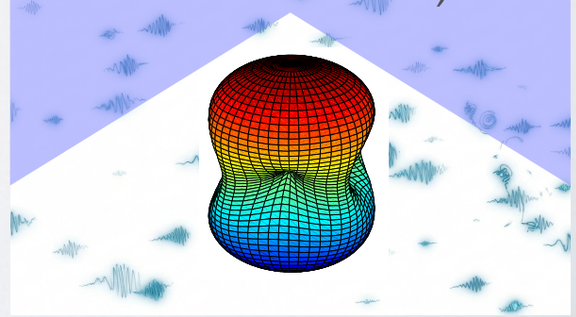
$$d(t) = n_{\mathrm{Easy}}(t) + n_{\mathrm{Hard}}(t) + R\left[\sum_i h_i(t, \vec{\lambda}_i)\right]$$

Detector noise that is easy to predict or model. Usually fundamental, and determines the detector sensitivity.



Detector/environment noise that is hard or impossible to predict or model. Usually technical, and determines the quality of the data.

Response to the superposition of all GW signals impinging at the time *t*, each with its own vector of parameters describing the source and possibly larger structures (even the whole Universe).



Detector Noise

- The combination of all the noise sources in a detector produces the discrete time series $\mathbf{n} = \{n_k\} = \{n(t_k)\}$

- Noise is described as a stochastic process with statistical properties given by the joint probability distribution $p(\mathbf{n})$

- Noise is Gaussian if this distribution follows a multi-variate normal distribution:

$$p(\mathbf{n}) = \frac{1}{\sqrt{\det 2\pi\mathbf{C}}} \exp\left[-\frac{1}{2}\sum_{ij}(n_i-\mu)C_{ij}^{-1}(n_j-\mu)\right]$$

where $\mu = E[\mathbf{n}]$ is the mean and $C_{ij} = E[(n_i-\mu)(n_j-\mu)]$ is the covariance matrix

Detector Noise

- Noise is stationary if the elements of the covariance matrix depend only on the time lag $|i-j|$ (and are constant on the diagonal)

- Switching to the Fourier domain the covariance matrix of stationary noise is diagonal: $C_{ij} = \delta_{ij} S_{\rm n}(f_i)$ (indices denote frequency bins; ~ over variables are omitted for simplicity)

- $S_{\rm n}(f_i)$ is the detector **power spectral density (PSD)**; it is not known a priori and must be estimated from the data (Welch averaging): it is a pivotal ingredient in data analysis

Detector Noise

- The idealized case of independent detectors with stationary and Gaussian noise is suitable for n_{Easy}

- Actual analyses carefully account for deviations from this

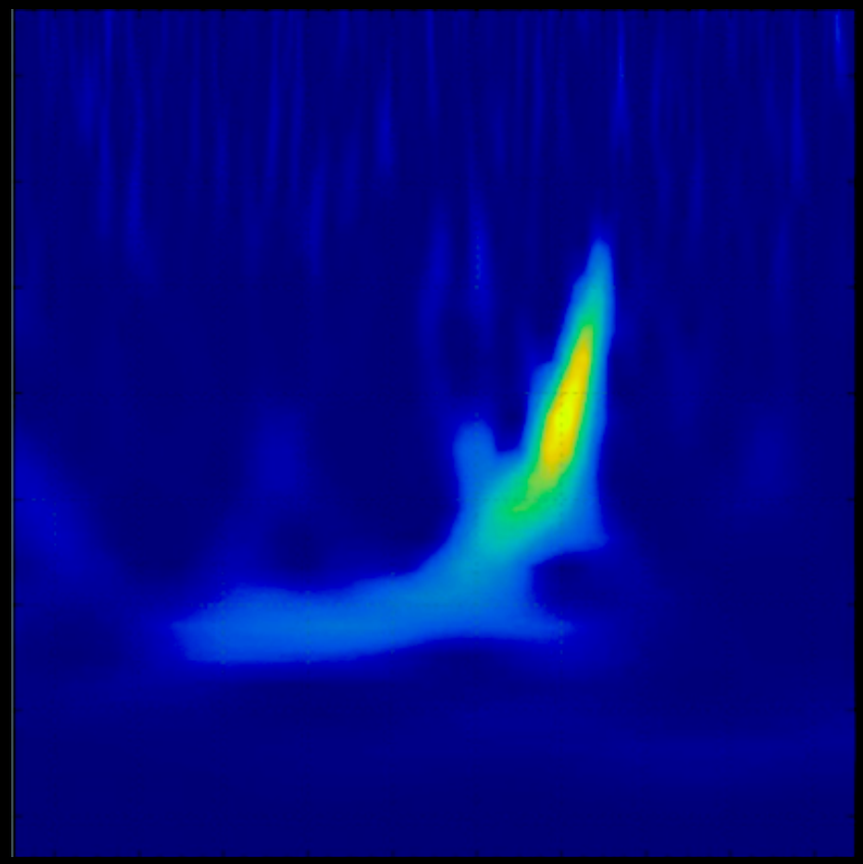
- Data exhibits two main types of non-stationary behaviour:

  1. slow and continuous adiabatic drifts in the power spectrum occurring over minutes, hours, days, or seasons

  2. short-duration noise transients, referred to as **glitches**, that are localized in time and frequency

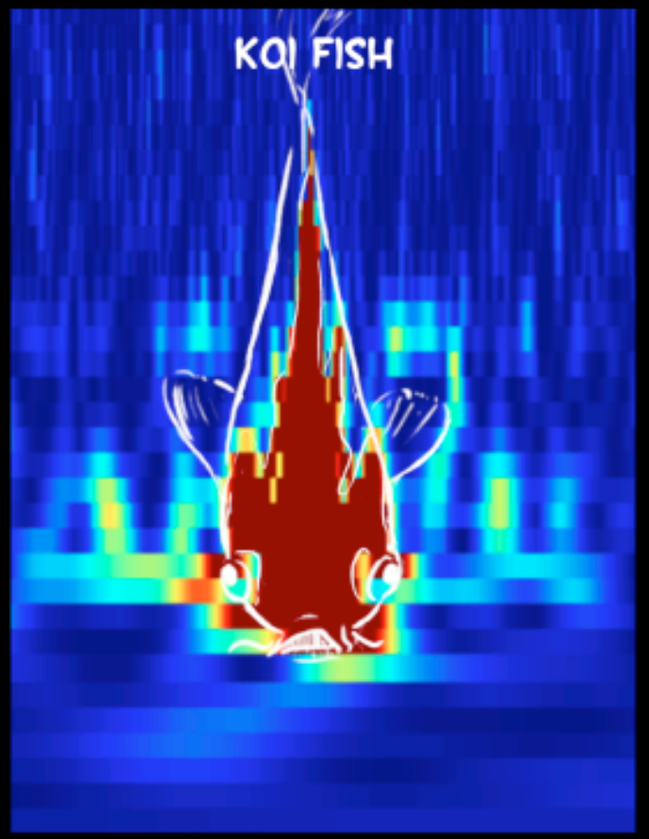
THE ART OF NAMING GLITCHES

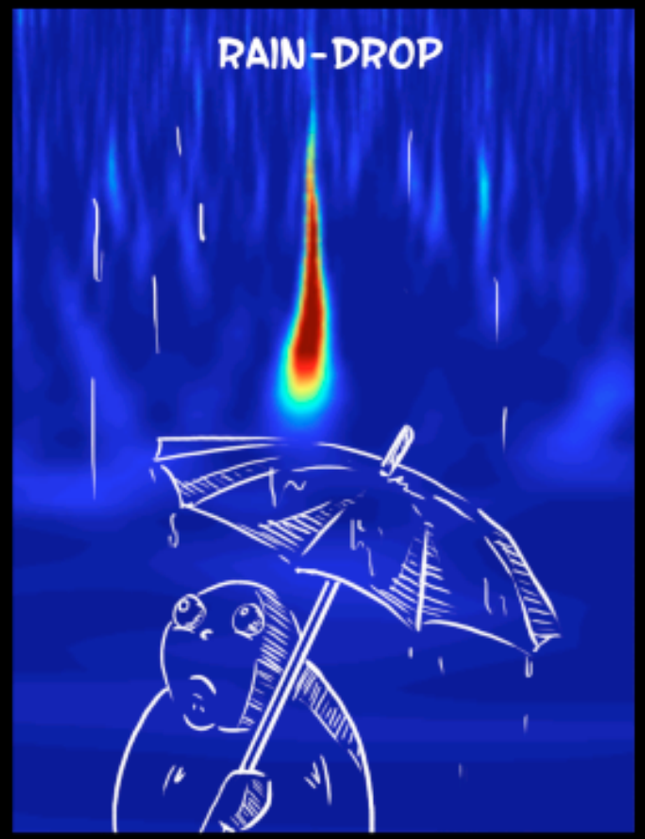
GW150914

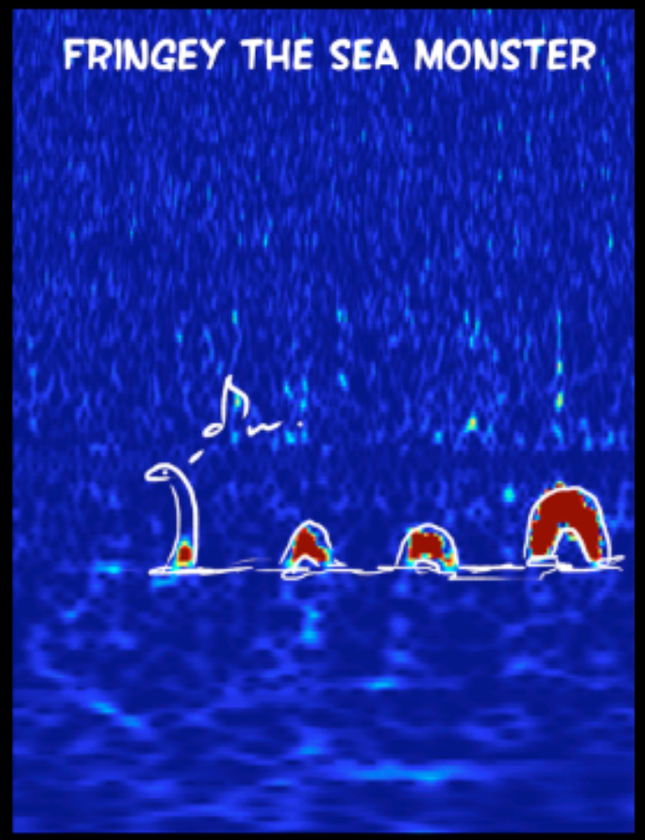


Frequency

Time







Detector Noise

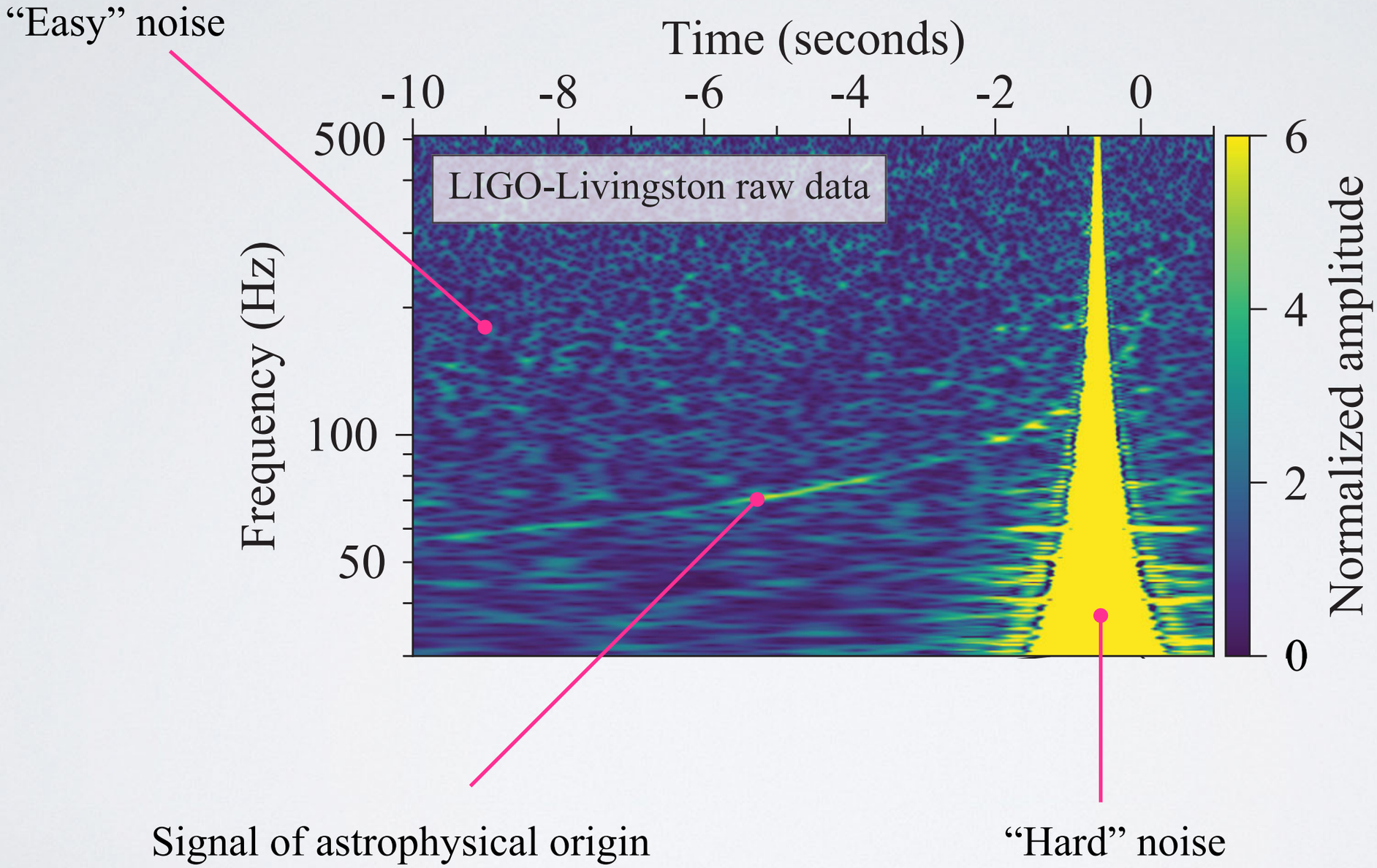
- Real detector noise is an arbitrary superposition of stationary, nonstationary, deterministic, nondeterministic, transient, persistent signals, the origins of which may be known or unknown

- Any component that can be understood and/or modeled can be subtracted from the data

- The rest can only be studied by running the detectors, observing how the noise behaves and trying to track down its origin in the detector (noise hunting)

- Distinct contributions to n_{Hard} may be more or less problematic for specific analyses

Detector Noise: Group Photo



GW Astronomy Building Blocks

Detector design and construction

Build the machines that provide $d(t)$

- Make n_{Easy} as small as possible
- Make $n_{\mathrm{Hard}} \ll n_{\mathrm{Easy}}$ so we can forget about it

Observing run

Collect $d(t)$

- Run the detector to obtain $d(t)$ for periods as long as possible
- Understand the detector behaviour and learn to improve it

Data analysis

Derive scientific knowledge from $d(t)$

- Identify each signal $h(t)$ (handling $n_{\mathrm{Easy}} + n_{\mathrm{Hard}}$ as best as possible)
- Characterize the source
- Use information from individually and multiple signals for science

Detect and Infer

(Frequentist) search

1. Gather the data (and deal with data quality, calibration, ...)
2. Build the foreground, i.e., identify candidates
   a. If using models, look through the data for your expectations
   b. Handle non-stationary (non-Gaussian) noise
3. Assess the signficance of candidates: build a background and compare it to the foreground

Bayesian inference

4. If a candidate qualifies as a signal, calculate Bayes Factor and infer source properties (otherwise place upper limits, make exclusion statements)

Astro-statistics

5. Infer (cosmological parameters, neutron star equation of state, population properties, etc.) from multiple signals

Two Key Ingredients

1. Bayes' theorem:

$$p(A \mid B)p(B) = p(B \mid A)p(A)$$

where $p(A) \in [0,1]$ is the probability of statement A

2. The likelihood function

Likelihood

- The likelihood that the gravitational-wave detector strain data contains a given signal is a central quantity for both searches and parameter estimation

- A gravitational wave hits the K-th detector, which records $\mathbf{d}_K = R_k[\mathbf{h}] + \mathbf{n}_K$, so at a single data point i we ask: are **noise** and **residual** probability compatible?

$$p(d_{K,i} - R_K[h'_i]) = p(n_{K,i})$$

- We do not know the signal exactly!
- Modelled and unmodelled searches depart here

- This is the probability of drawing the former argument from our noise model, under the null hypothesis $\mathcal{H}_0$

Likelihood

- The likelihood function is the noise model (hence all those slides on noise…)

- We have discrete frequency bins and want the joint probability for the noise from all frequency bins:

$$p(\mathbf{d}_K|\mathbf{h}') = p(d_{K,1} - R_K[h'_1], \ldots, d_{K,N_f} - R_K[h'_{N_f}])$$
$$= p(n_{K,1}, \ldots, n_{K,N_f})$$

- Likelihood for Gaussian noise

$$p(\mathbf{d}|\mathbf{h}') = p(\mathbf{n}) = \frac{1}{\sqrt{\det 2\pi\mathbf{C}}} \exp\left[-\frac{1}{2}\sum_{(IJ),(km)} (d_{I,k} - R_I[h'_k]) C^{-1}_{(Ik)(Jm)} (d_{J,m} - R_J[h'_m])\right]$$

where we sum over detectors (IJ) and data samples (km)

Likelihood

- If the detectors are uncorrelated: $C_{(Ik),(Jm)} = \delta_{IJ} S^I_{km}$

[block diagonal]

- If they are also stationary: $C_{(Ik),(Jm)} = \delta_{IJ}\delta_{km} S^I(f_k)$

[diagonal]

- In this case, modulo an overall negative sign, the argument of the exponential in the noise likelihood reads

$$\frac{1}{2}\chi^2(\mathbf{d},\mathbf{h}') = \frac{1}{2}\langle \mathbf{d}-\mathbf{h}'|\mathbf{d}-\mathbf{h}'\rangle = \frac{1}{2}\sum_K \langle \mathbf{d}_K - R_K[\mathbf{h}']|\mathbf{d}_K - R_K[\mathbf{h}']\rangle$$

$$\xrightarrow{\text{continuum}} 2\Re \sum_K \int_0^\infty \frac{(d_K(f) - R_K[h'(f)])^*(d_K(f) - R_K[h'(f)])}{S^K(f)} df$$

Dropping the prime for a lighter notation

Detect and Infer

(Frequentist) search

1. Gather the data (and deal with data quality, calibration, ...)
2. Build the foreground, i.e., identify candidates
   a. If using models, look through the data for your expectations
   b. Handle non-stationary (non-Gaussian) noise
3. Assess the signficance of candidates: build a background and compare it to the foreground

Bayesian inference

4. If a candidate qualifies as a signal, calculate Bayes Factor and infer source properties (otherwise place upper limits, make exclusion statements)

Astro-statistics

5. Infer (cosmological parameters, neutron star equation of state, population properties, etc.) from multiple signals

Signal Detection

- Searches compare null hypothesis and signal hypothesis

- From Bayes' theorem, the **posterior probability** of the signal hypothesis given the observed data is

$$p(\mathcal{H}_1|\mathbf{d}) = p(\mathbf{d}|\mathcal{H}_1)p(\mathcal{H}_1)/p(\mathbf{d}) = \frac{p(\mathbf{d}|\mathcal{H}_1)p(\mathcal{H}_1)}{p(\mathbf{d}|\mathcal{H}_0)p(\mathcal{H}_0) + p(\mathbf{d}|\mathcal{H}_1)p(\mathcal{H}_1)} =$$

$$= \frac{p(\mathbf{d}|\mathcal{H}_1)}{p(\mathbf{d}|\mathcal{H}_0)} \left[\frac{p(\mathbf{d}|\mathcal{H}_1)}{p(\mathbf{d}|\mathcal{H}_0)} + \frac{p(\mathcal{H}_0)}{p(\mathcal{H}_1)}\right]^{-1}$$

which is monotonic in the **likelihood ratio** of the two hypotheses

$$\Lambda(\mathbf{d}|\mathbf{h}(\vec{\lambda})) = \frac{p(\mathbf{d}|\mathcal{H}_1)}{p(\mathbf{d}|\mathcal{H}_0)}$$

Signal Detection

- The (log-)likelihood function can be used to define a frequentist detection statistic given by the likelihood ratio between a signal being present or absent in the data

$$\log \Lambda(\mathbf{d}|\mathbf{h}(\vec{\lambda})) = p(\mathbf{d}|\mathcal{H}_1) - p(\mathbf{d}|\mathcal{H}_0)$$

- For stationary and Gaussian noise this statistic would follow a known distribution and the false alarm rate for an event could be computed analytically

$$\log \Lambda(\mathbf{d}|\mathbf{h}(\vec{\lambda})) = -\frac{1}{2}\langle \mathbf{d} - \mathbf{h}(\vec{\lambda})|\mathbf{d} - \mathbf{h}(\vec{\lambda})\rangle + \frac{1}{2}\langle \mathbf{d}|\mathbf{d}\rangle =$$

$$= \langle \mathbf{h}(\vec{\lambda})|\mathbf{d}\rangle - \frac{1}{2}\langle \mathbf{h}(\vec{\lambda})|\mathbf{h}(\vec{\lambda})\rangle$$

Residual

Matched filter

"Scale factors" that do not mix $\mathbf{h}$ and $\mathbf{d}$

Signal Detection

- The posterior probability is a monotonic function of the matched filter term which is therefore an optimal test statistic

- Maximising the likelihood to search for signals is equivalent to:
  - maximising the matched filter or the **signal-to-noise ratio** $\sqrt{\langle \mathbf{h}(\vec{\lambda}) \,|\, \mathbf{d} \rangle / \langle \mathbf{h}(\vec{\lambda}) \,|\, \mathbf{h}(\vec{\lambda}) \rangle}$
  - minimising the residuals

- In practice noise exhibits deviations from stationarity and Gaussianity: robust search methods have been developed to take into account the measured properties of the noise and deal with these deviations

Search Families and Approaches

- **Transient** maximum likelihood searches look for times when the likelihood (or possibly the matched filter, or the signal-to-noise ratio) peaks

- **Persistent** maximum likelihood searches consider as much data as possible at once and "accumulate likelihood"

Search Families and Approaches

- **Incoherent** transient searches tackle the likelihood maximisation at each detector:
  - peaks above a detection statistic threshold are promoted to coincident triggers if they are present in all detectors compatibly with the light travel time between them
  - they are ranked with the quadrature sum of the signal-to-noise ratio in each detector (or variations of this definition)

- **Coherent** searches address the likelihood maximisation simultaneously in all detectors with a common signal:
  - requires exploring in sky location (larger parameter space)
  - operate directly on the $h_+, h_\times$ polarisations
  - formally more solid, and beneficial for many detectors

Search Families and Approaches

- **Modelled** maximum likelihood searches
  - assume a waveform model, which is a "recipe" to calculate $\mathbf{h}(\vec{\lambda})$ given the signal parameters $\vec{\lambda}$
  - maximise the likelihood by processing $\mathbf{d}$ with a finite number of predetermined $\vec{\lambda}$ choices (templates)

- **Unmodelled** maximum likelihood searches
  - do not assume a specific signal morphology, but use for example a flexible decomposition in wavelets
  - perform the maximisation of the likelihood ratio (typically coherently) over the possible sky position obtaining $h_+, h_\times$ as a solution

Search Families and Approaches

- **Cross-correlation** searches are based on the idea that cross-correlating data among (pairs of or groups of) detectors averages out noise fluctuations but not astrophysical signals
  - rather than maximise the likelihood, they essentially average away noise
  - strong hypothesis is that correlated noise is negligible

Detect and Infer

(Frequentist) search

1. Gather the data (and deal with data quality, calibration, ...)
2. Build the foreground, i.e., identify candidates
   a. If using models, look through the data for your expectations
   b. Handle non-stationary (non-Gaussian) noise
3. Assess the signficance of candidates: build a background and compare it to the foreground

Bayesian inference

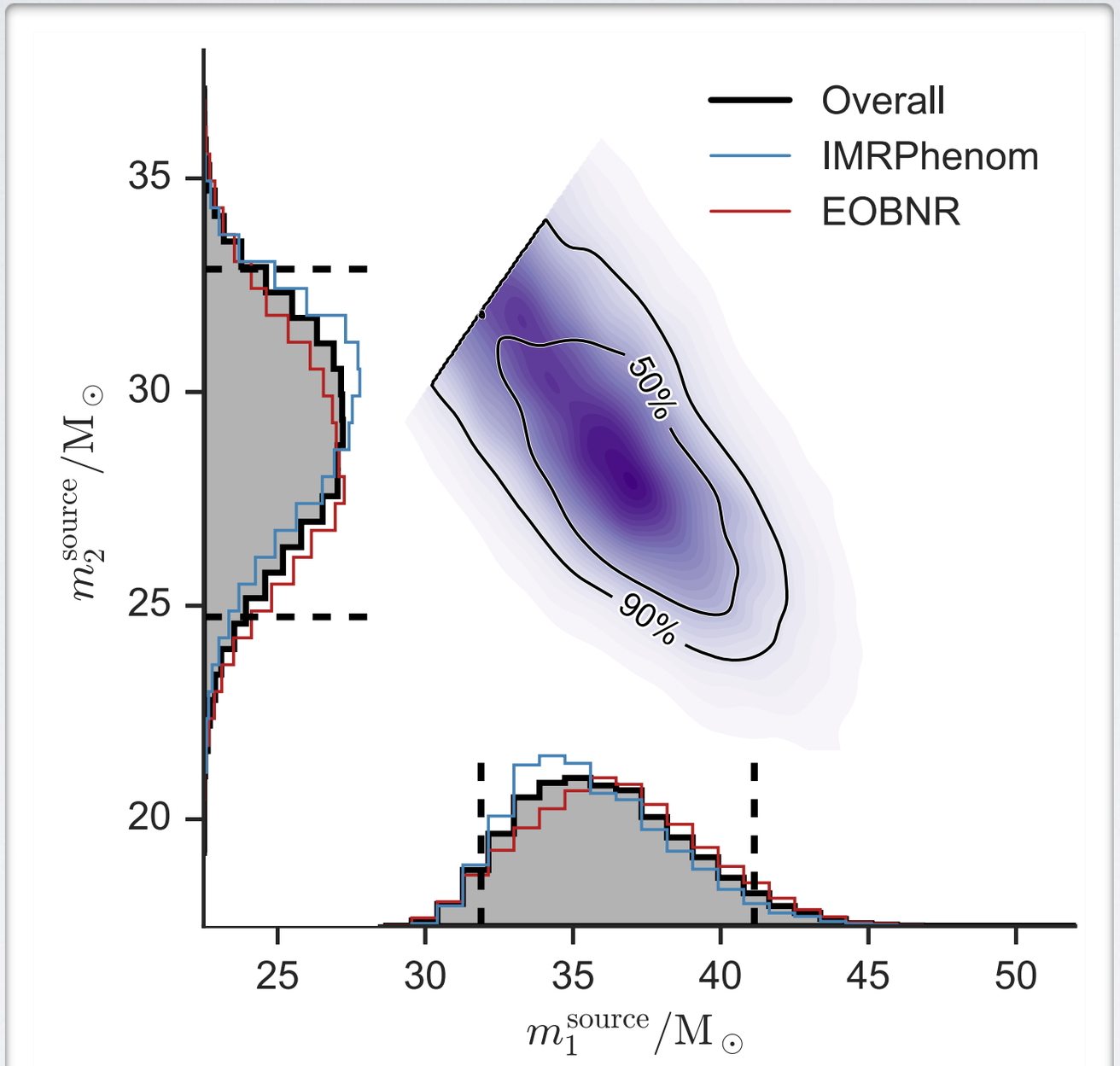
4. If a candidate qualifies as a signal, calculate Bayes Factor and infer source properties (otherwise place upper limits, make exclusion statements)

Astro-statistics

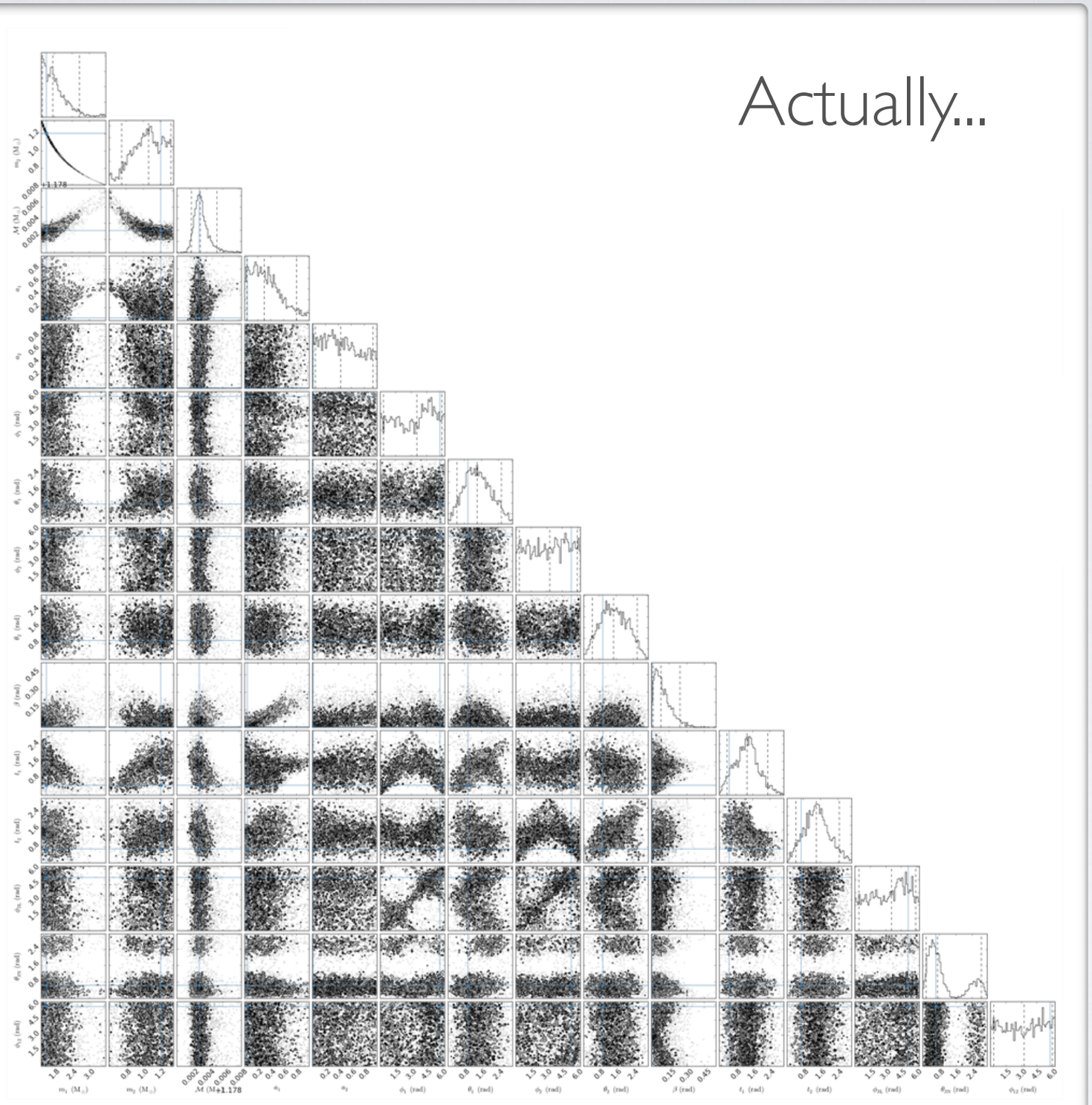
5. Infer (cosmological parameters, neutron star equation of state, population properties, etc.) from multiple signals

Gravitational-Wave Inference

Going from the detection to…



Gravitational-Wave Inference



Actually...

Bayes' Theorem: Take 2

$$p(A \mid B)p(B) = p(B \mid A)p(A)$$

$$p(\vec{\lambda}|\mathbf{d}, \mathcal{M}_h, \mathcal{I}) = \frac{p(\mathbf{d}|\vec{\lambda}, \mathcal{M}_h, \mathcal{I})\ p(\vec{\lambda}|\mathcal{M}_h, \mathcal{I})}{p(\mathbf{d}|\mathcal{M}_h, \mathcal{I})}$$

Bayes' Theorem: Take 2

Updated understanding = new observations + initial understanding

Likelihood

Prior

$$p(\vec{\lambda}|\mathbf{d},\mathcal{M}_h,\mathcal{I}) = \frac{p(\mathbf{d}|\vec{\lambda},\mathcal{M}_h,\mathcal{I})\,p(\vec{\lambda}|\mathcal{M}_h,\mathcal{I})}{p(\mathbf{d}|\mathcal{M}_h,\mathcal{I})}$$

Posterior

Evidence

$\vec{\lambda}$ in the case of compact binaries: masses (2), spins (6), sky position (2), distance (1), orbital plane inclination (1), polarization (1), phase and time at coalescence (2), finite size effects, deviations from General Relativity, ...

Posterior

- The posterior gives the probability density that a model $h(\vec{\lambda})$ describes the data

$$p(\vec{\lambda}|\mathbf{d}, \mathcal{M}_h, \mathcal{I}) = \frac{p(\mathbf{d}|\vec{\lambda}, \mathcal{M}_h, \mathcal{I})\ p(\vec{\lambda}|\mathcal{M}_h, \mathcal{I})}{p(\mathbf{d}|\mathcal{M}_h, \mathcal{I})}$$

Posterior

- It is calculated with a likelihood and a prior, and it is valid under the assumptions that were used when computing them

Likelihood

- The likelihood function is central to Bayesian inference: with the specification of priors for the signal and noise models, it allows for the calculation of the model evidence – odds that a signal is present – and posterior distributions for the model

Likelihood

$$p(\vec{\lambda}|\mathbf{d},\mathcal{M}_h,\mathcal{I}) = \frac{p(\mathbf{d}|\vec{\lambda},\mathcal{M}_h,\mathcal{I})\,p(\vec{\lambda}|\mathcal{M}_h,\mathcal{I})}{p(\mathbf{d}|\mathcal{M}_h,\mathcal{I})}$$

$$p(\mathbf{d}|\vec{\lambda},\mathcal{M}_h,\mathcal{I}) \equiv \Lambda(\mathbf{d}|\mathbf{h}(\vec{\lambda})) = \frac{p(\mathbf{d}|\mathcal{H}_1)}{p(\mathbf{d}|\mathcal{H}_0)}$$

$$p(\mathbf{d}|\mathcal{M}_h,\mathcal{I}) = \int p(\mathbf{d}|\vec{\lambda},\mathcal{M}_h,\mathcal{I})p(\vec{\lambda}|\mathcal{M}_h,\mathcal{I})d\vec{\lambda}$$

Evidence

Dropping $\mathcal{I}$ for a lighter notation

Likelihood

- Remember the two critical assumptions (at each detector):
  - Gaussian noise $p(n_{K,1},\ldots,n_{K,N_f}) \propto \exp\left[-\frac{1}{2}n_i C^{-1}_{K,ij} n_j\right]$
  - Stationary noise $C_{K,ij} = \frac{1}{2}S_K(f_i)\delta_{ij}$
- Probability of obtaining data assuming signal hypothesis and that the noise is Gaussian and stationary

Noise-weighting

$$p(\mathbf{d}|\vec{\lambda},\mathcal{M}_h) \propto \exp\left[-\frac{1}{2}\sum_K \langle d_K - R_K[h(\vec{\lambda})]|d_K - R_K[h(\vec{\lambda})]\rangle\right]$$

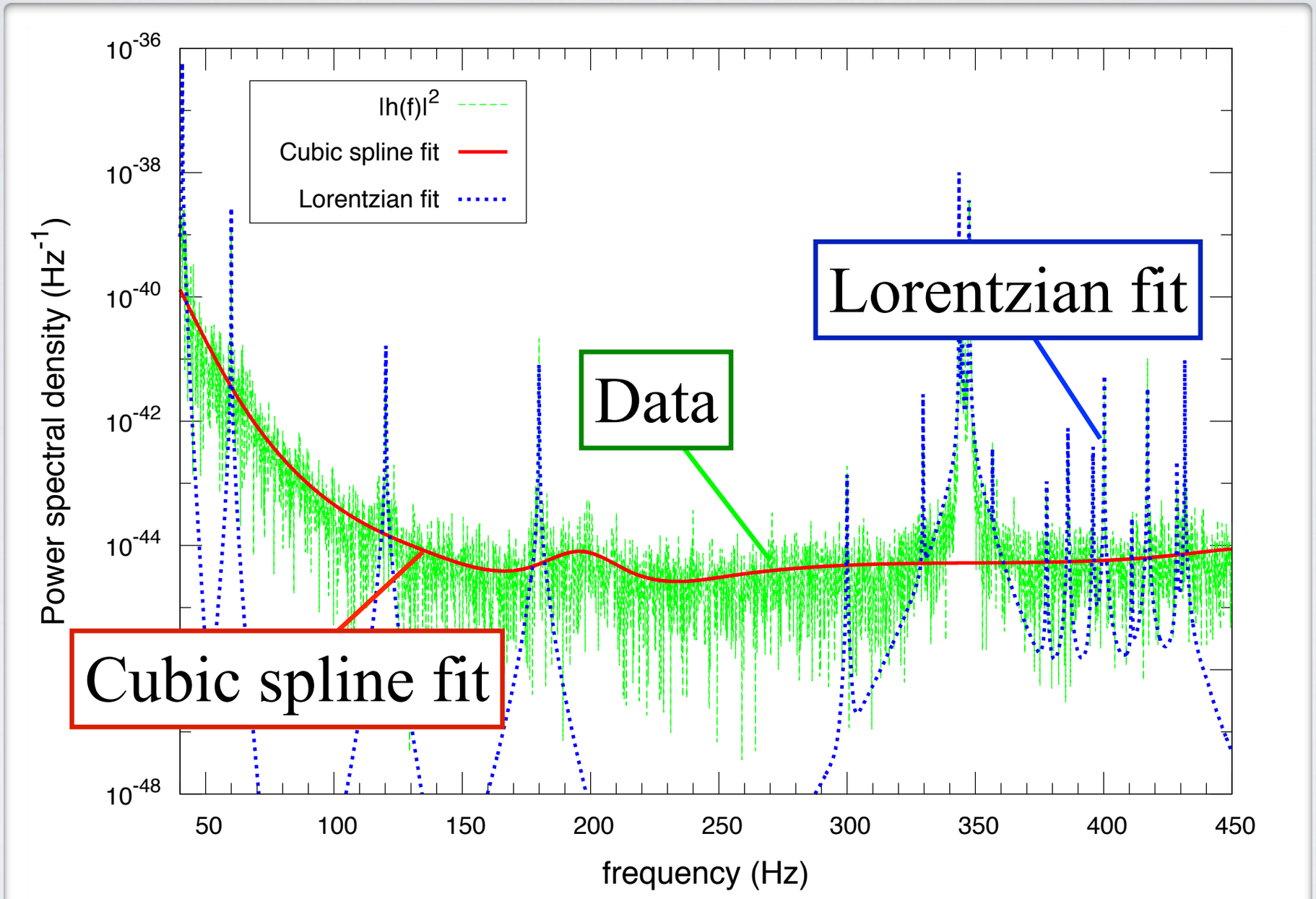
$$h_i(\vec{\lambda}) \rightarrow h'_i(\vec{\lambda})(1+\delta A_i)\exp(i\delta\phi_i)$$

Waveform model

Calibration
O(10) parameters
per detector

Handling Non-Stationarity

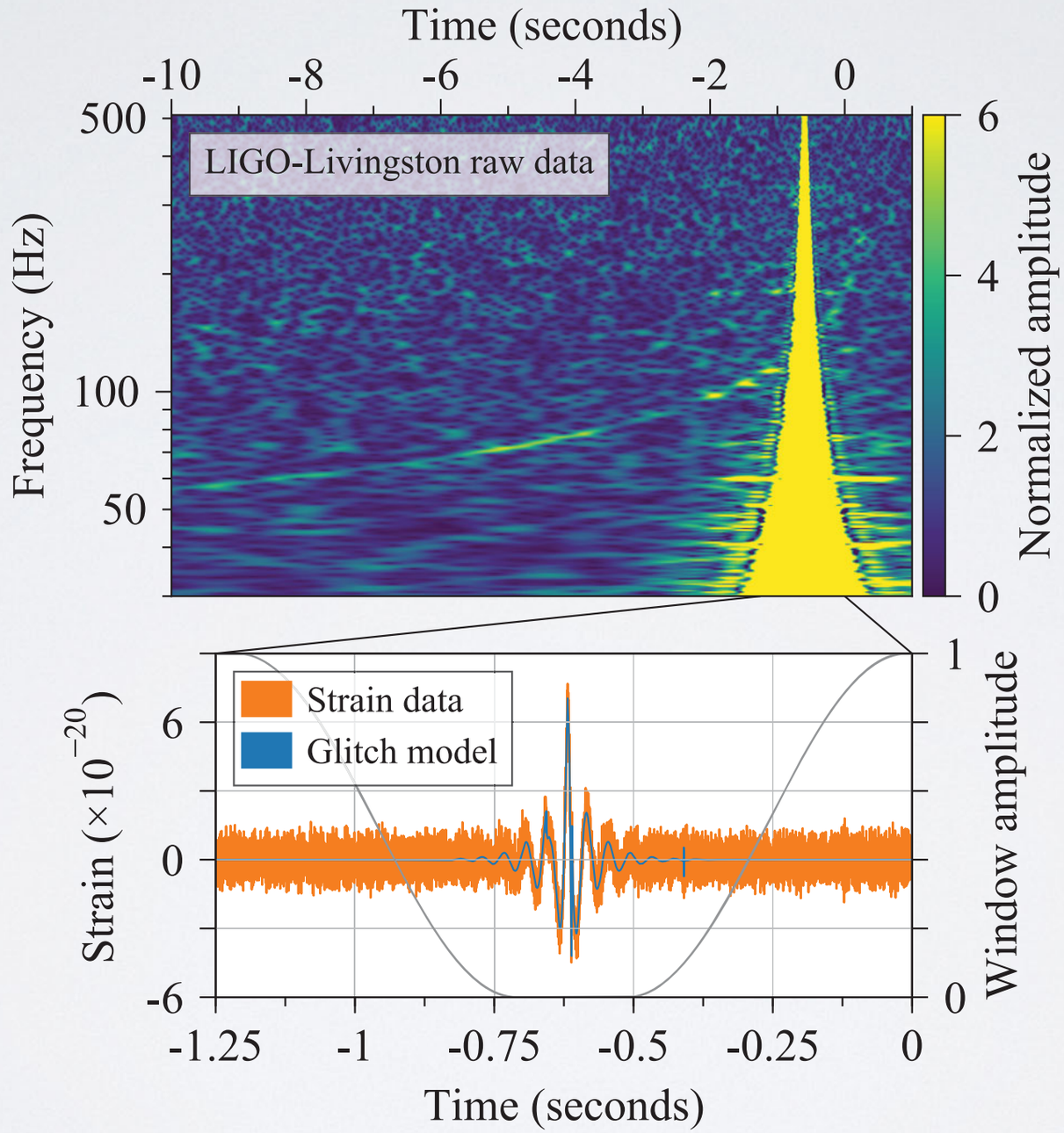
Recompute the PSD for every detector and for every event



[BayesLine – arXiv:1410.3852]

Handling Non-Gaussianity

Include glitch in likelihood ($\mathbf{d - R[h] - g = n}$) or remove it from data



[GW170817 – arXiv:1710.05832]

Gravitational-Wave Inference

$$p\left(\tilde{\mathbf{d}}\middle|\vec{\lambda},\mathcal{M}_h\right) = \exp\left[-\frac{1}{2}\left(\tilde{\mathbf{d}}-\tilde{\mathbf{h}}(\vec{\lambda})\middle|\tilde{\mathbf{d}}-\tilde{\mathbf{h}}(\vec{\lambda})\right)\right]$$

"likelihood"

data *waveform*

$$\text{"Inner product"} = \int \frac{\left|\tilde{d}(f)-\tilde{h}(f;\vec{\lambda})\right|^2}{S_n(f)}df$$

noise PSD

$$p\left(\vec{\lambda}\middle|\mathcal{M}_h\right)$$

"prior"

"posterior"

$$p\left(\vec{\lambda}\middle|\tilde{\mathbf{d}},\mathcal{M}_h\right)$$

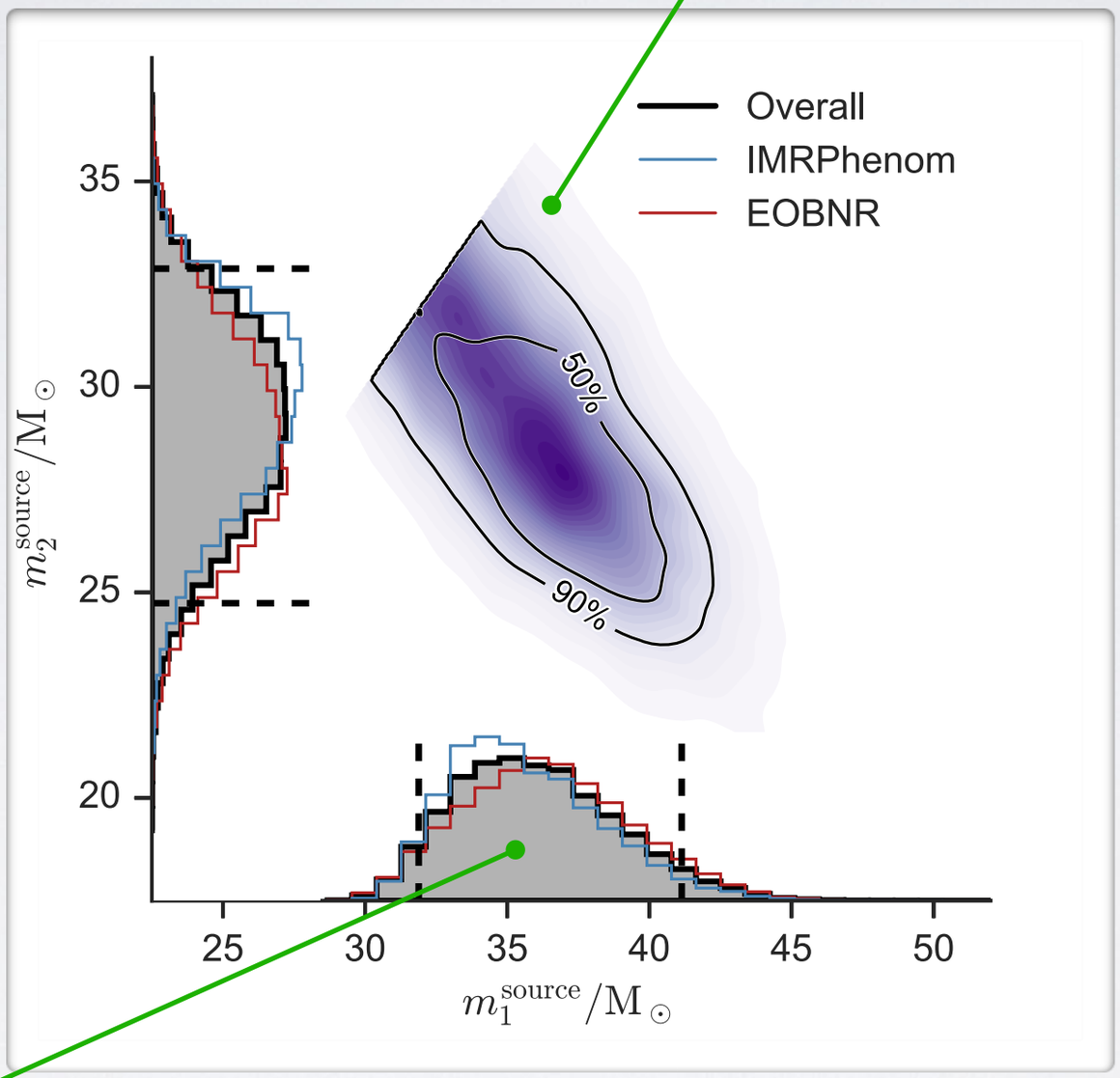
"evidence"

$$p\left(\tilde{\mathbf{d}}\middle|\mathcal{M}_h\right)$$

[Credit: T.Littenberg]

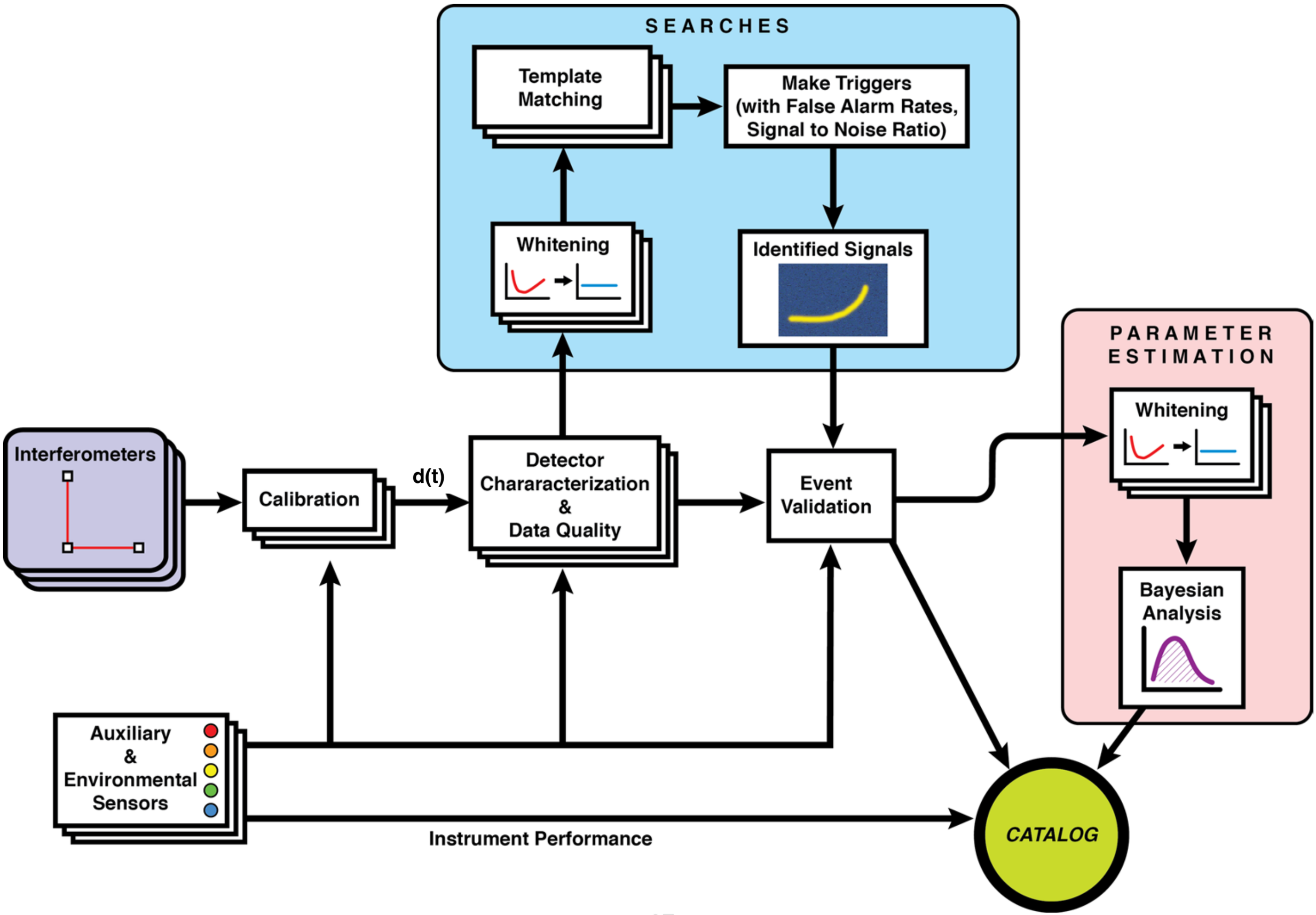
Marginalized Posteriors

$$p(\lambda_1, \lambda_2|\mathbf{d}, \mathcal{M}_h) = \int p(\vec{\lambda}|\mathbf{d}, \mathcal{M}_h) d\lambda_3 \dots d\lambda_N$$



$$p(\lambda_1|\mathbf{d}, \mathcal{M}_h) = \int p(\vec{\lambda}|\mathbf{d}, \mathcal{M}_h) d\lambda_2 \dots d\lambda_N$$

[arXiv:1602.03840]


SEARCHES
Template Matching
Make Triggers (with False Alarm Rates, Signal to Noise Ratio)
Whitening
Identified Signals
PARAMETER ESTIMATION
Whitening
Bayesian Analysis
Interferometers
Calibration
d(t)
Detector Chararacterization & Data Quality
Event Validation
Auxiliary & Environmental Sensors
Instrument Performance
CATALOG

Additional Resources

- LIGO-Virgo Collaboration, "A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals" Class. Quantum Grav. **37**, 055002 (2020)
- Finn, "Detection, measurement and gravitational radiation" PhRvD **46**, 5236 (1992)
- Ashton et al., "Bilby: A user-friendly Bayesian inference library for gravitational-wave astronomy" ApJS **241**, 27 (2019)
- Creighton and Anderson, "Gravitational-Wave Physics and Astronomy: An Introduction to Theory, Experiment and Data Analysis" Wiley-VCH Verlag GmbH & Co. KGaA (2011)
- Thrane and Talbot, "An introduction to Bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models" Publications of the Astronomical Society of Australia **36**, e010 (2019)
- Pitkin, Messenger, Fan, "Hierarchical Bayesian method for detecting continuous gravitational waves from an ensemble of pulsars" PhRvD **98**, 063001 (2018)
- https://gwosc.org/
- https://www.zooniverse.org/projects/zooniverse/gravity-spy