The statistics of ground-based interferometric gravitational-wave detection and astrophysics

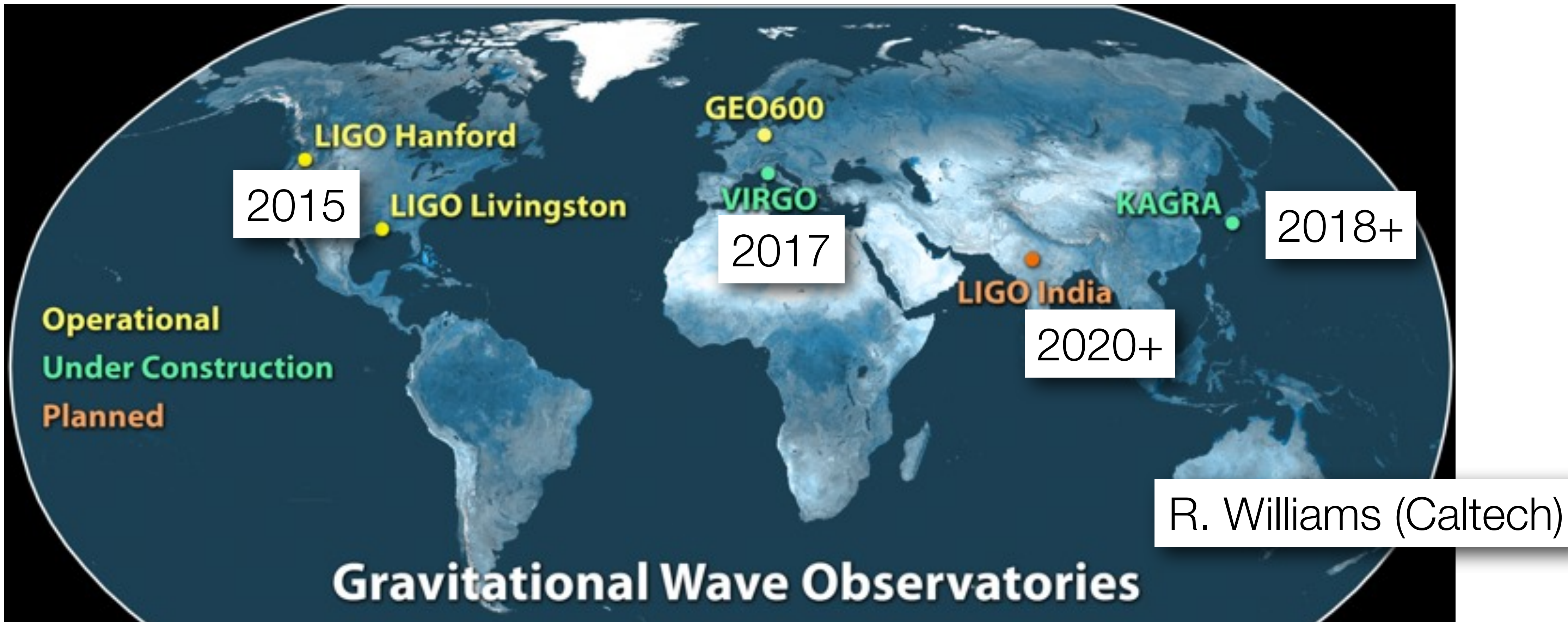
Chris Pankow / CIERA / Northwestern University
PhyStat-Nu, September 19, 2016

Special thanks: Will Farr (Birmingham), Kipp Cannon (Tokyo), Jolien Creighton (UW-Milwaukee), Neil Cornish (Montana State), Tyson Littenberg (NASA/MSFC) Meg Millhouse (Montana State)

LIGO-G1601914

The LVC: Who We Are

- **The LIGO and Virgo Collaborations**: 1000+ scientists, engineers, and others spread amongst 50+ academic institutions world wide (presence on all continents except Africa and Antarctica)



- Collectively develop and operate a network of three kilometer-scale interferometers (LIGO Hanford, LIGO Livingston, Virgo), and a 600m pathfinder interferometer (GEO600)

- Two kilometer-scale interferometers under construction (KAGRA collaboration, Japan) or in design process (LIGO India)

Gravitational-Wave Source Spectrum

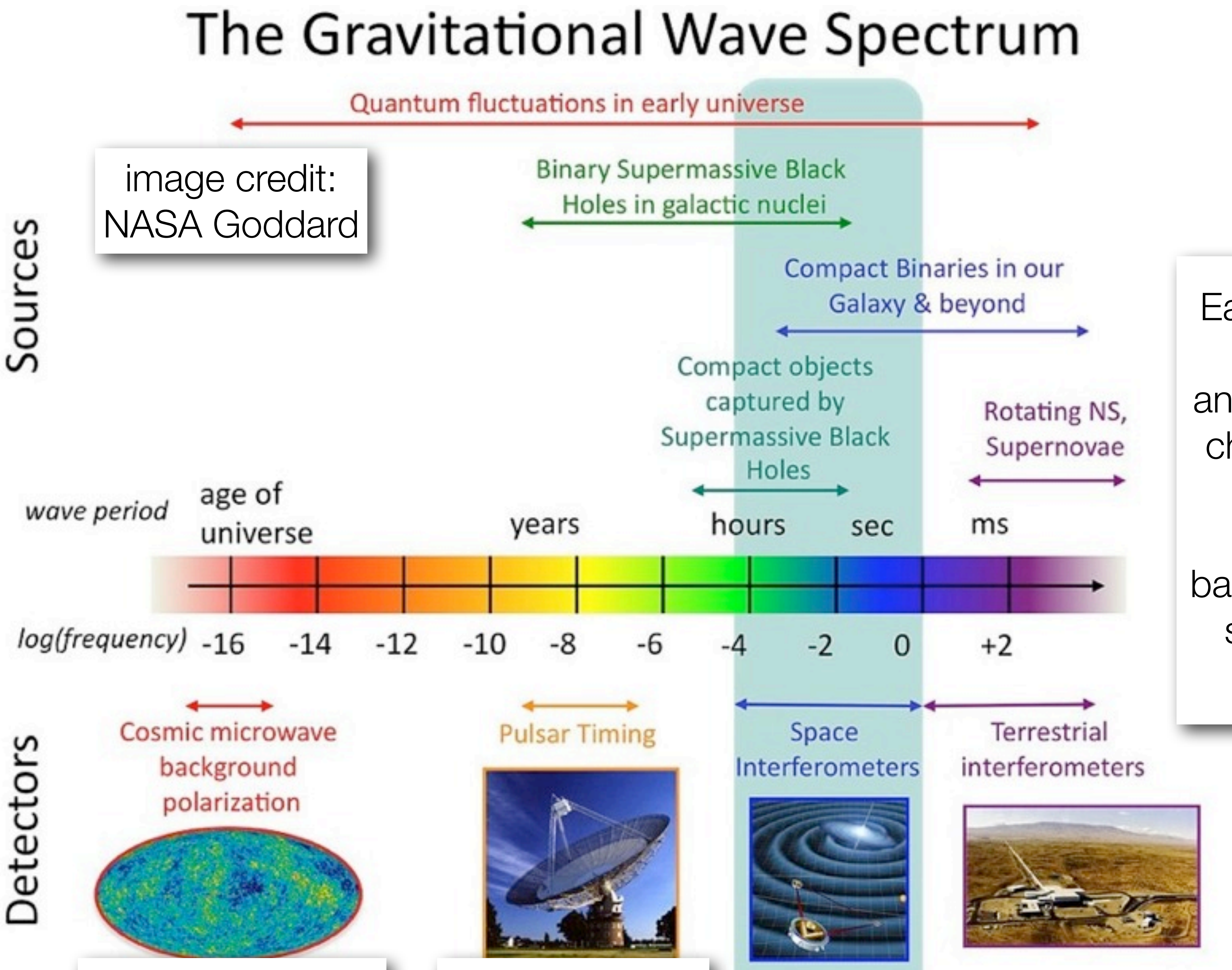


image credit:
NASA Goddard

Each frequency band presents different analysis and statistical challenges: unevenly sampled data, complicated background modeling, source foreground confusion, etc...

BICEP, PolarBear, etc..

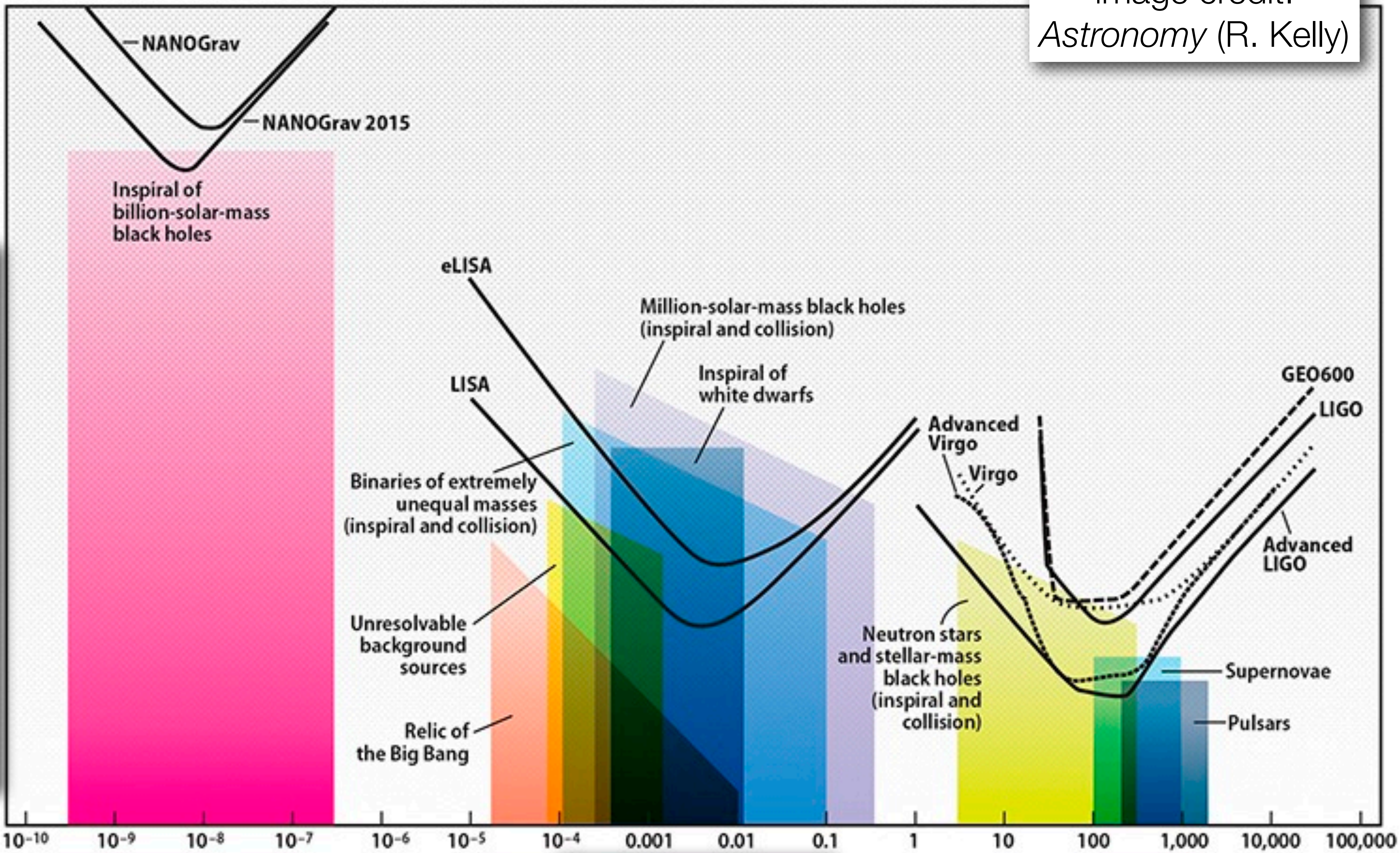
NANOGrav, Parkes, etc..

(e)LISA

LIGO / Virgo

Gravitational-Wave Source Spectrum

image credit: *Astronomy* (R. Kelly)

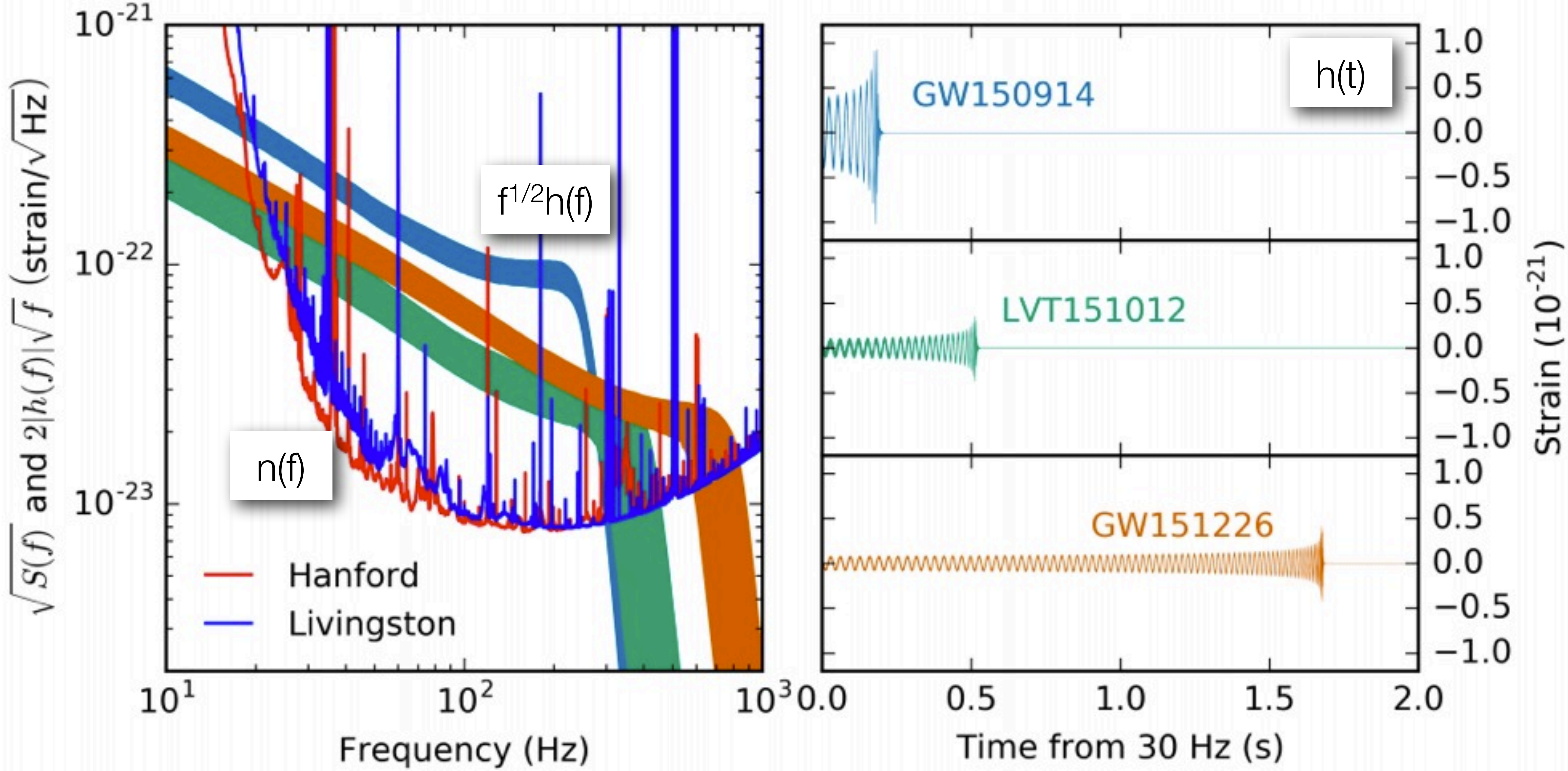


strain sensitivity (logarithmic, arb.)

frequency (Hz)

O1 BBH Events

arxiv:1606.04856



Gravitational-wave detection and parameterization: Unique meld of "time domain" astronomy and spectral methods

BBH: "Chirps" in the time domain (monotonically increasing in frequency vs time)
Lower mass → Higher frequency content / longer "in band"

Basic Terminology

$$d(t) = n(t) + h(t)$$

observations: Putative strain from **gravitational wave** is embedded in **detector noise**

$$S(|f|) = 2\delta(f - f')\langle \tilde{n}(f)\tilde{n}(f')\rangle$$

Noise power spectrum: Autocorrelation of the noise in the frequency domain — "limiting factor" of the sensitivity of the instrument

$$(a|b) \equiv 2\int_{\infty}^{\infty} \frac{\tilde{a}^{\star}(f)\tilde{b}(f)}{S(f)}$$

Noise weighted inner product: frequency-domain cross-correlation between two quantities

Null Hypothesis (H_0): Data samples are uncorrelated Gaussian noise with variance proportional to S(f)

$$p(H_0) \propto \exp(-(d|d)/2)$$

Alternative Hypothesis (H_1): data are distributed as in null, *after* subtraction of the signal model (h)

$$p(H_1) \propto \exp(-(d-h|d-h)/2)$$

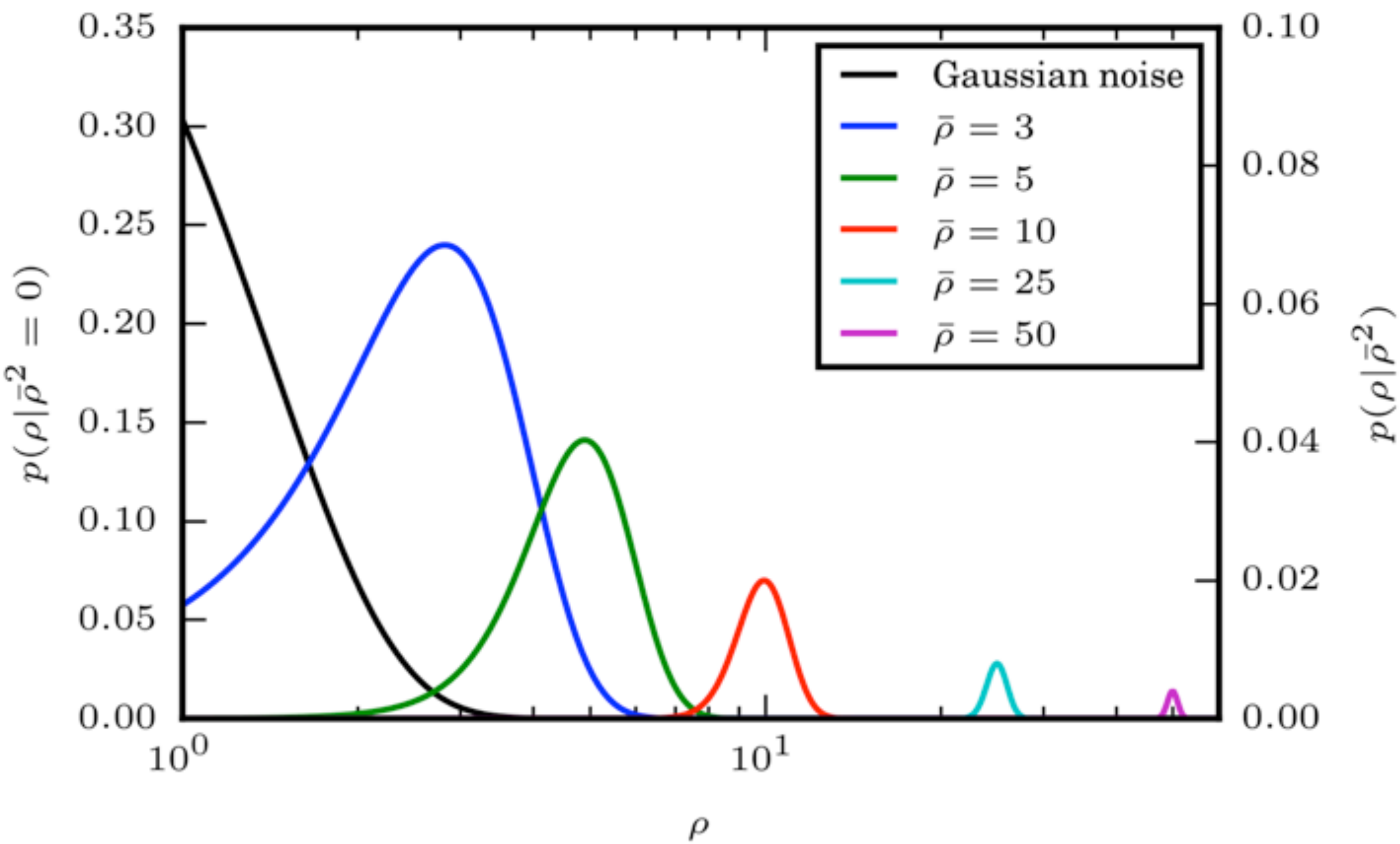
Likelihood Ratio / Signal-to-Noise Ratio

$$\Lambda = \frac{p(d|H_1)}{p(d|H_0)} = \exp\left(-\underbrace{(d|h)}_{=\rho} + \underbrace{(h|h)}_{=\bar{\rho}^2}/2\right)$$

"matched filter" SNR

"characteristic" SNR

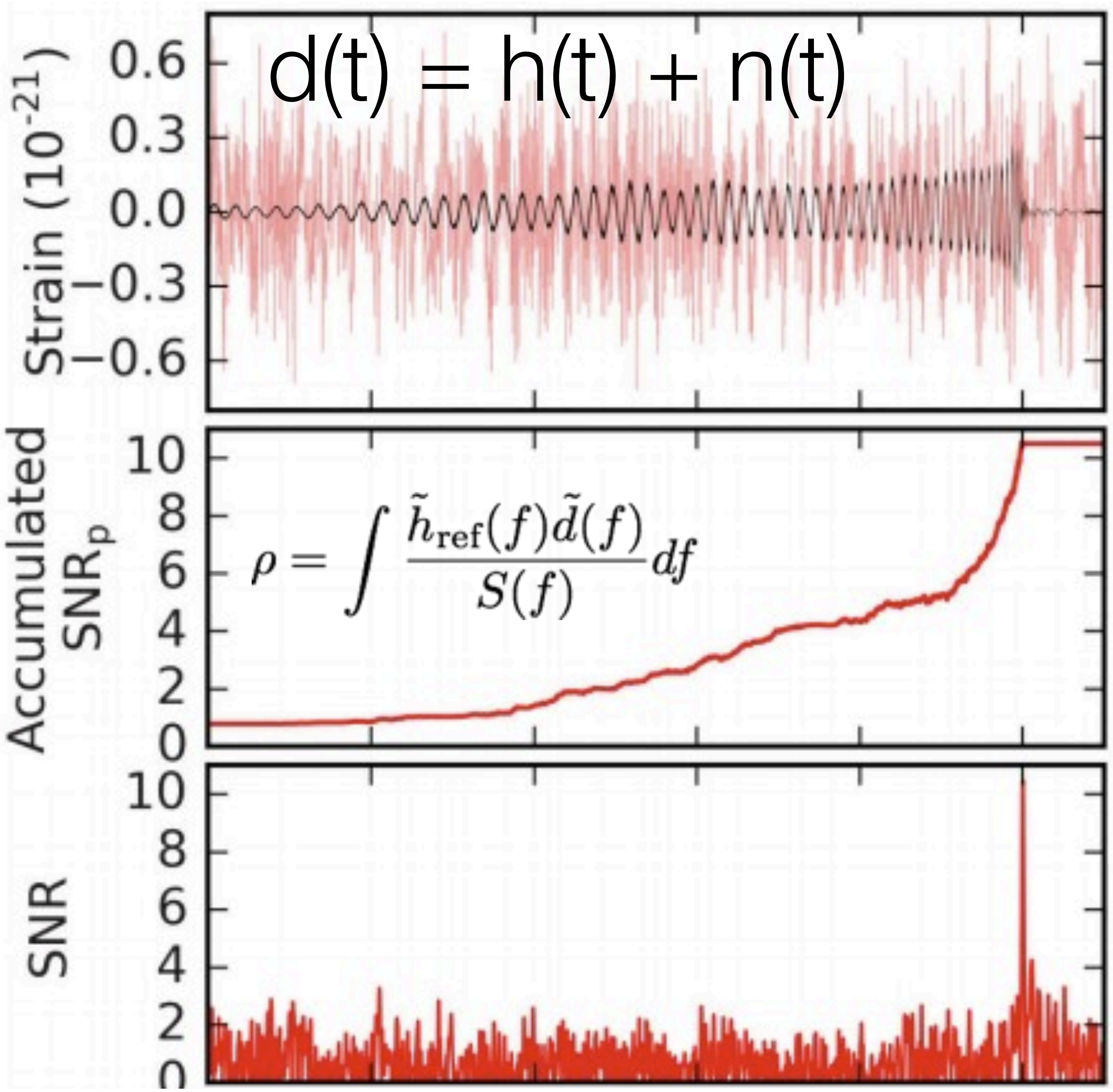
Form the "likelihood ratio": ratio of probability of signal present vs. probability of not present

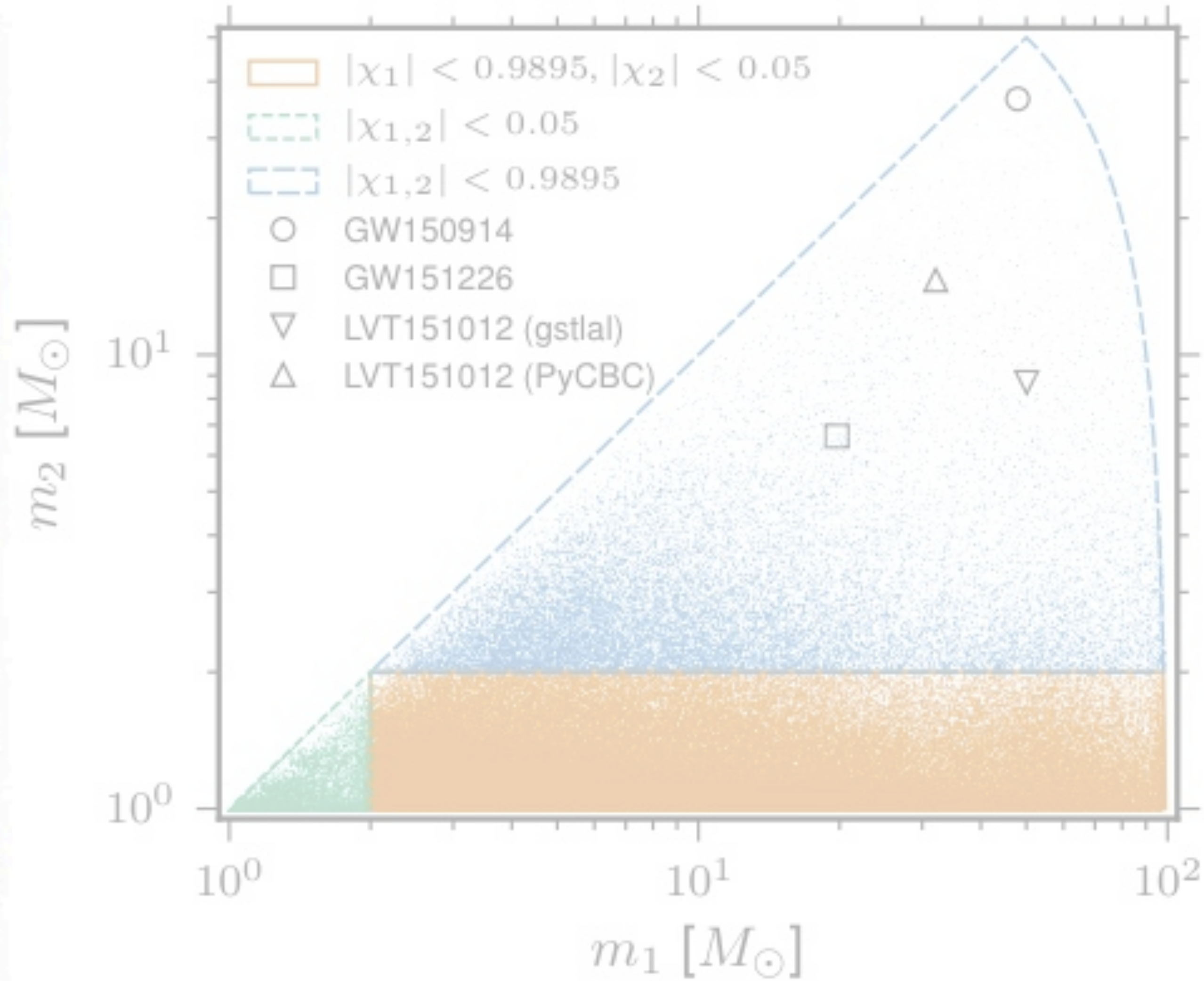


Invoke Neyman-Pearson lemma: At a given threshold, this is the *most powerful* test we can apply — maximizing the signal-to-noise ratio **ρ**=(d|h) maximizes the likelihood ratio

$\bar{\rho}$: What we *expect* (with perfect models)
ρ: The statistic we measure

GW Signal Detection Primer



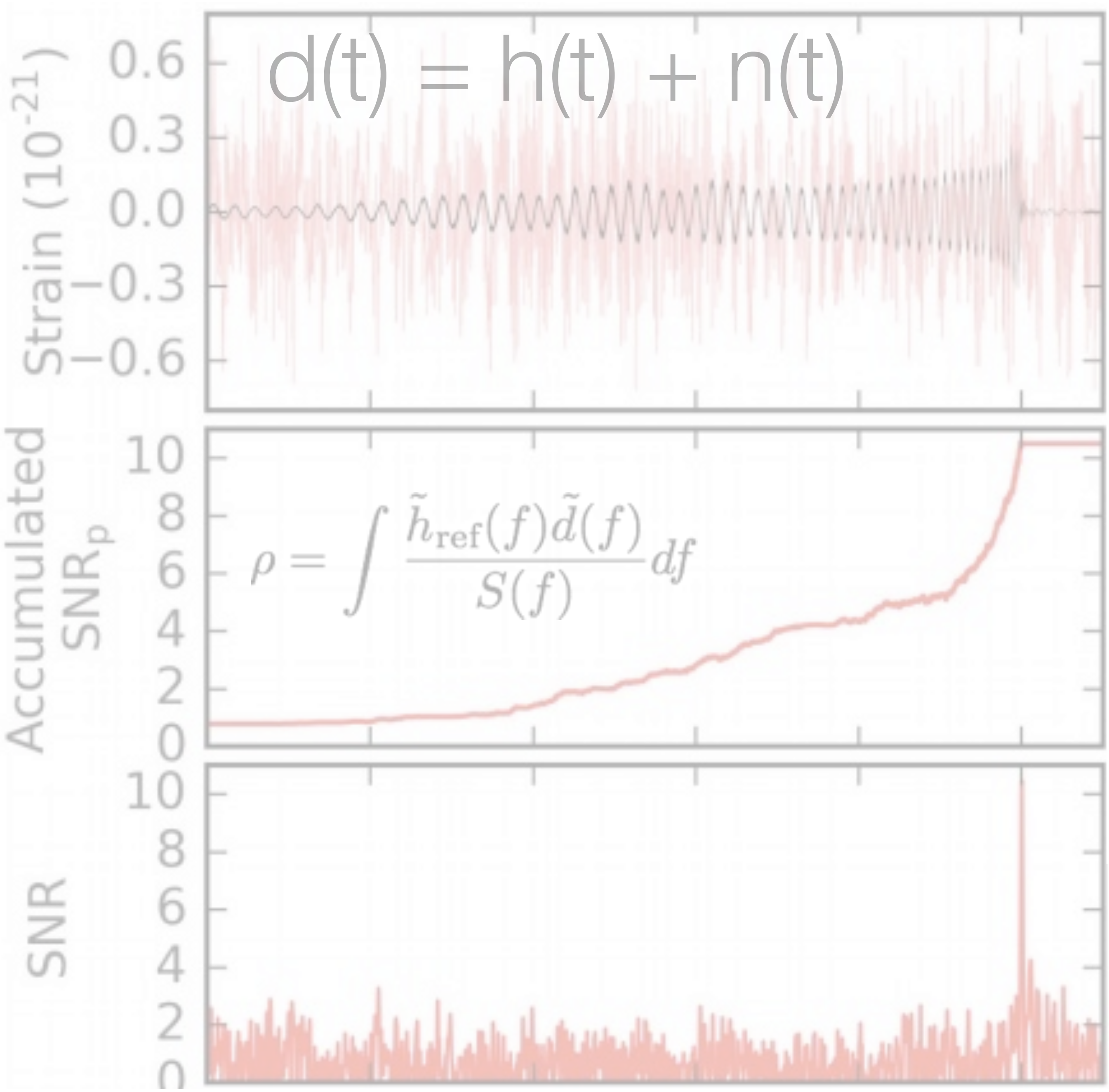


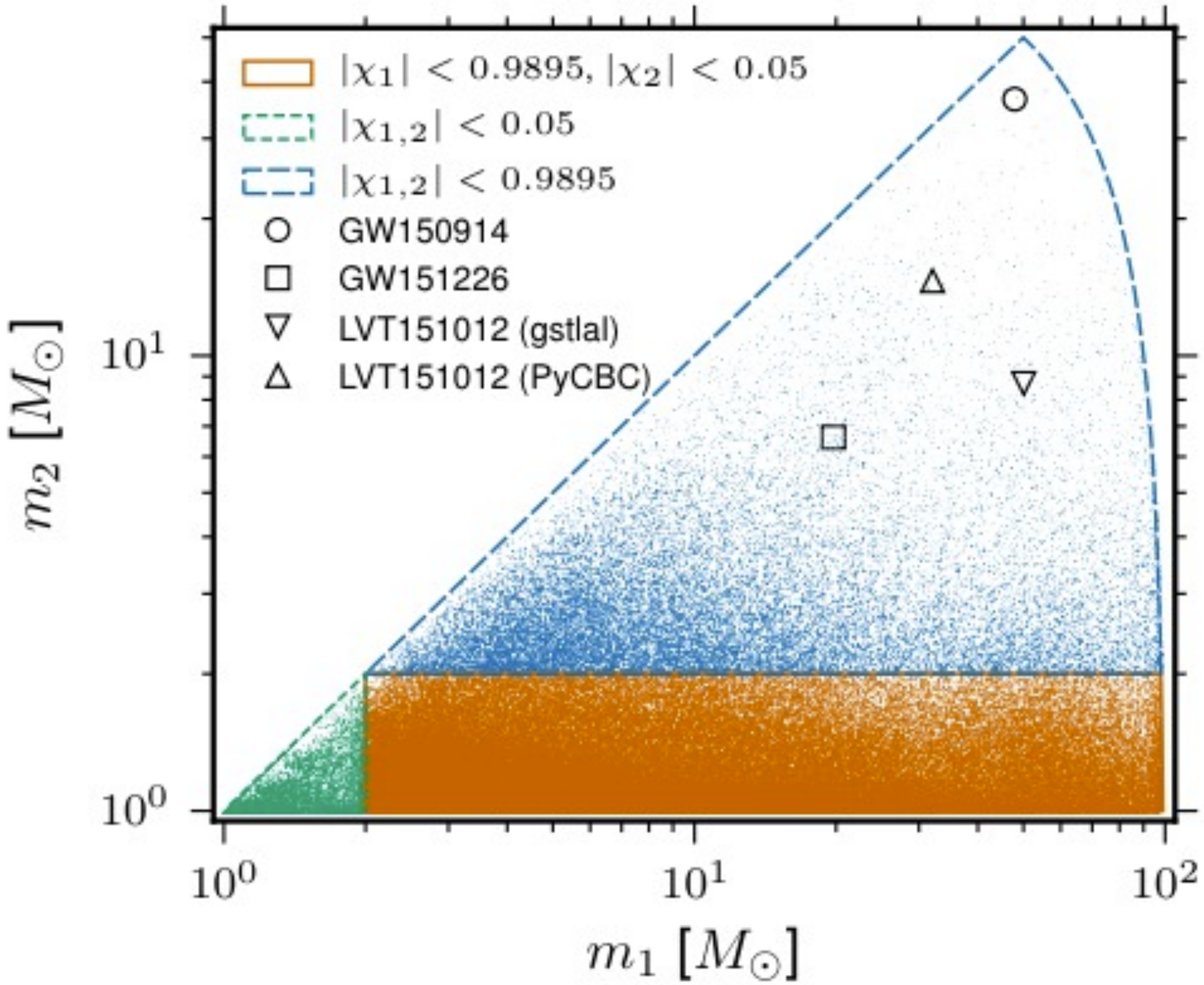
Putative strain is embedded in detector noise — cross correlate the model with the data to extract a signal-to-noise ratio (SNR, ρ) statistic — this maximizes the likelihood (probability of signal vs probability of noise)

Searches maximize likelihood analytically for speed and over masses/spins by brute force (template banks)

arxiv:1606.04856

GW Signal Detection Primer





Searches maximize likelihood analytically for speed and over masses/spins by brute force (template banks)

Putative strain is embedded in detector noise — cross correlate the model with the data to extract a signal-to-noise ratio (SNR, ρ) statistic — this maximizes the likelihood (probability of signal vs probability of noise)

arxiv:1606.04856

Bayesian Parameter Estimation

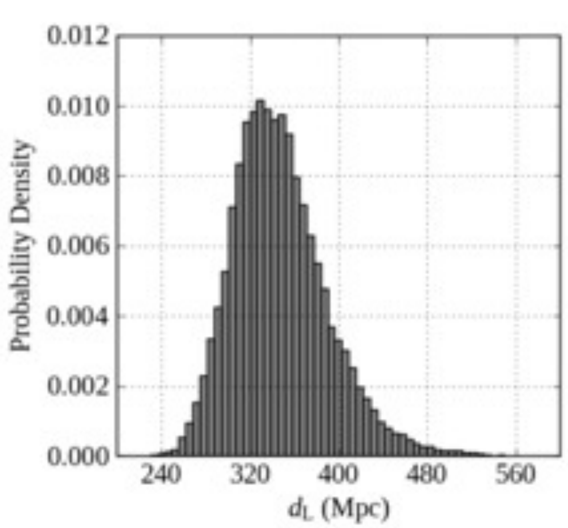
$$p(\mu|H_1,d) = \frac{p(\mu)\Lambda(d|\mu,H_1)}{p(d)}$$

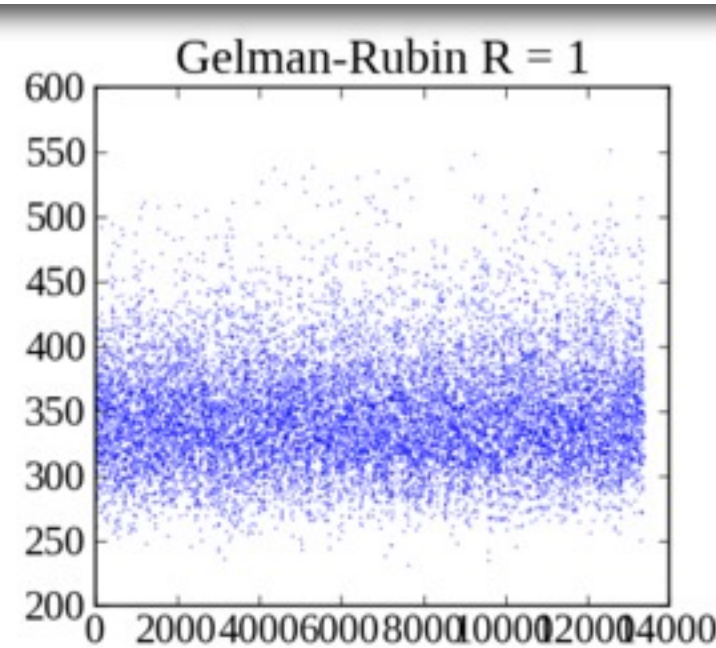
Parameter Posteriors: Form the *posterior* on a given parameter set μ from Bayes' Law

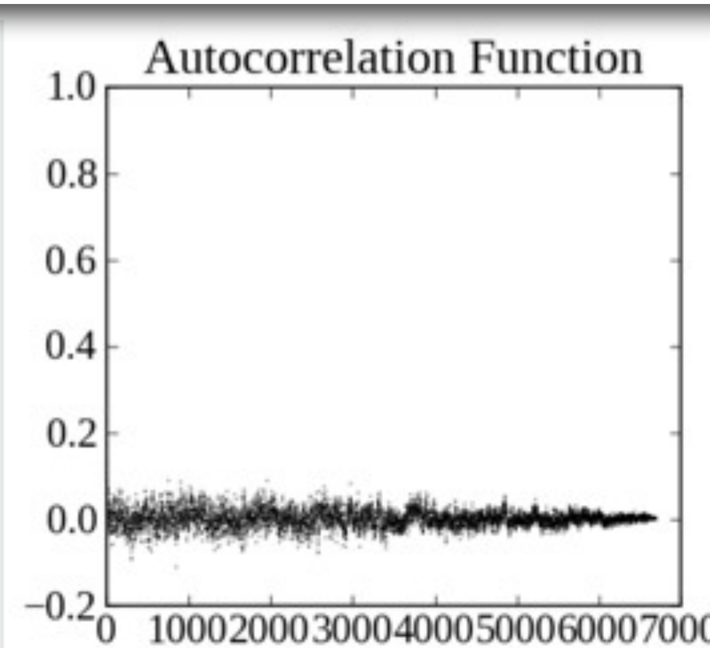
$$p(d) = \int p(\mu)\Lambda(d|\mu,H_1)d\mu$$

Bayes Factor: Often overlooked (posterior distributions normalized manually) but encodes the Bayesian signal vs. noise comparison

PE Method 1: 15+ dimensional space explored by Markov-Chain Monte Carlo. Parallel tempering: Λ^T added for more efficient convergence time. Determine "convergence" by number of effective, uncorrelated samples drawn using the autocorrelation length







Bayesian Parameter Estimation

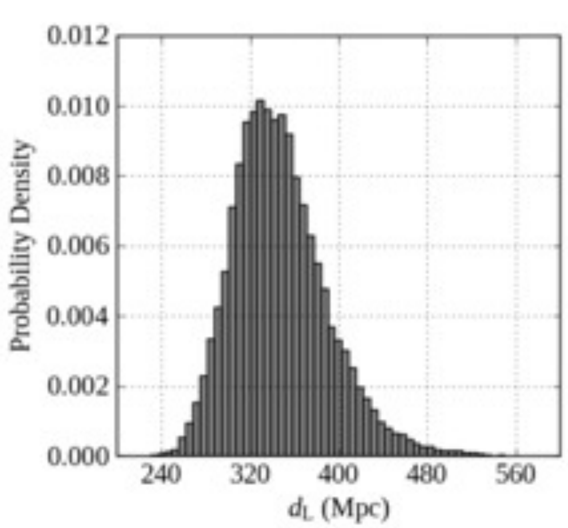
$$p(\mu|H_1, d) = \frac{p(\mu)\Lambda(d|\mu, H_1)}{p(d)}$$

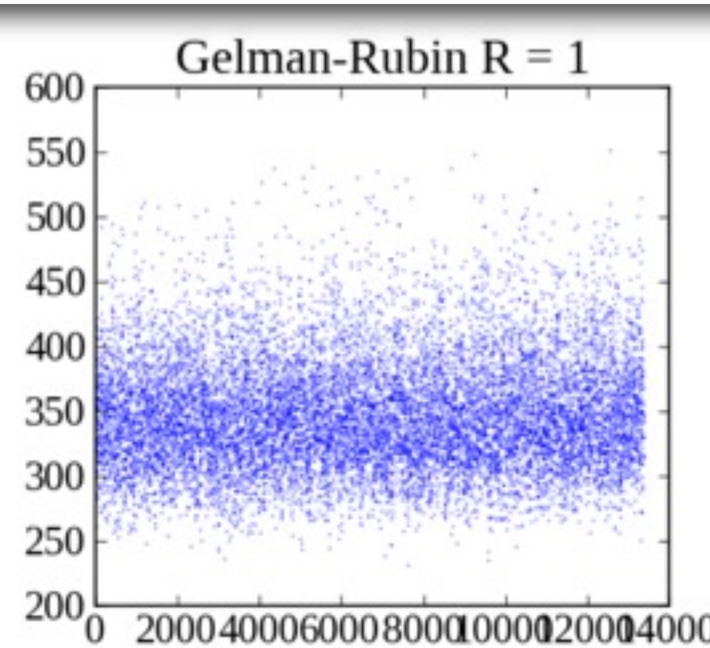
Parameter Posteriors: Form the *posterior* on a given parameter set μ from Bayes' Law

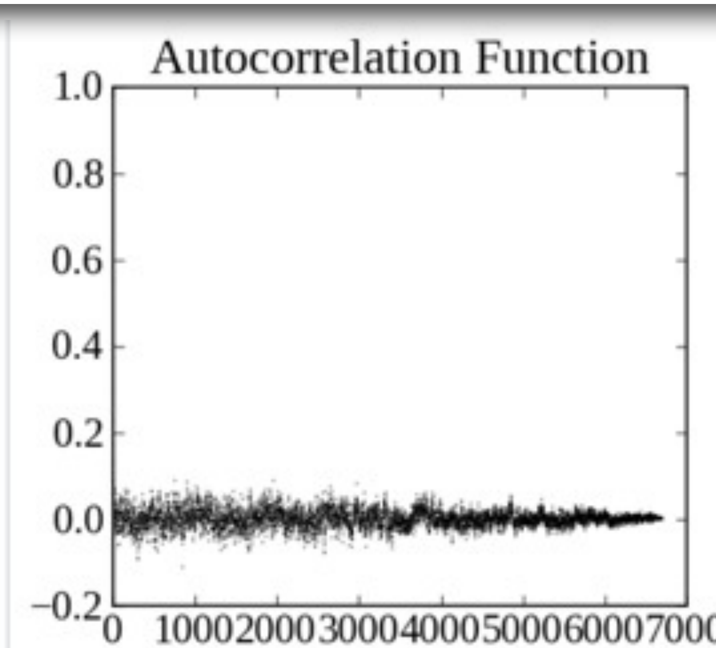
$$p(d) = \int p(\mu)\Lambda(d|\mu, H_1)d\mu$$

Bayes Factor: Often overlooked (posterior distributions normalized manually) but encodes the Bayesian signal vs. noise comparison

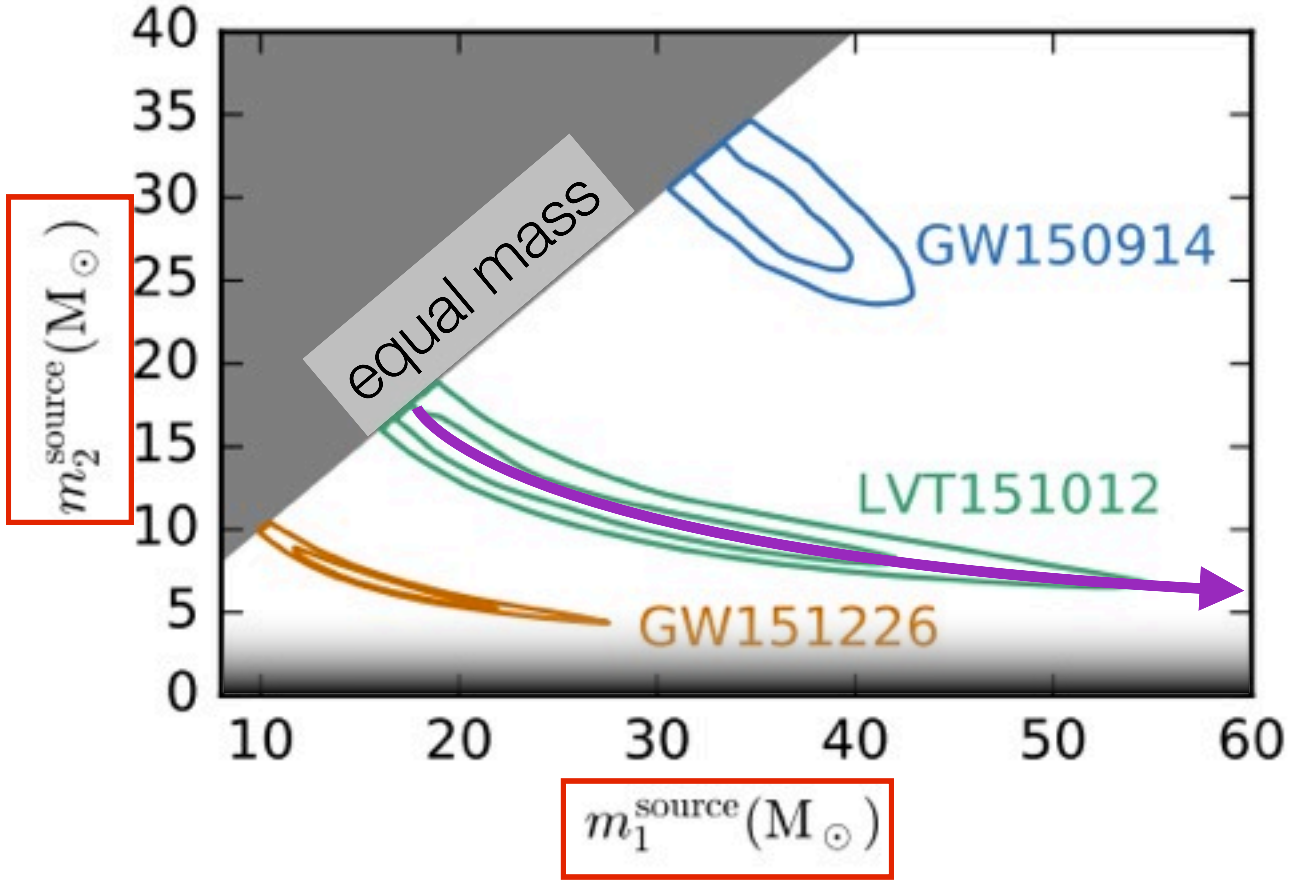
PE Method 2: "Nested Sampling"; swarm of points exploring the likelihood space and estimating the integrand. Allows much better estimation of the evidence (compared to thermodynamic integration for MCMC)







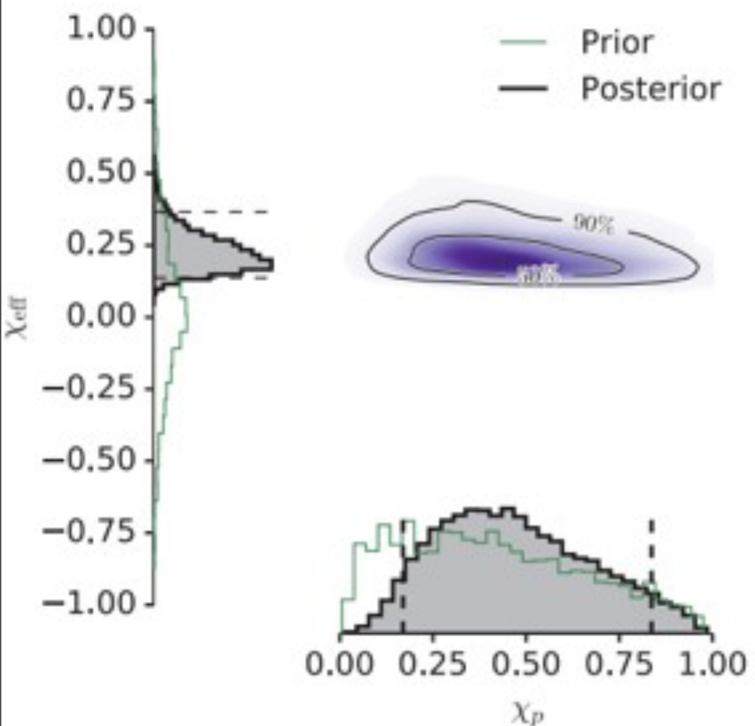
MCMC Param. Correlations: Masses and Spins

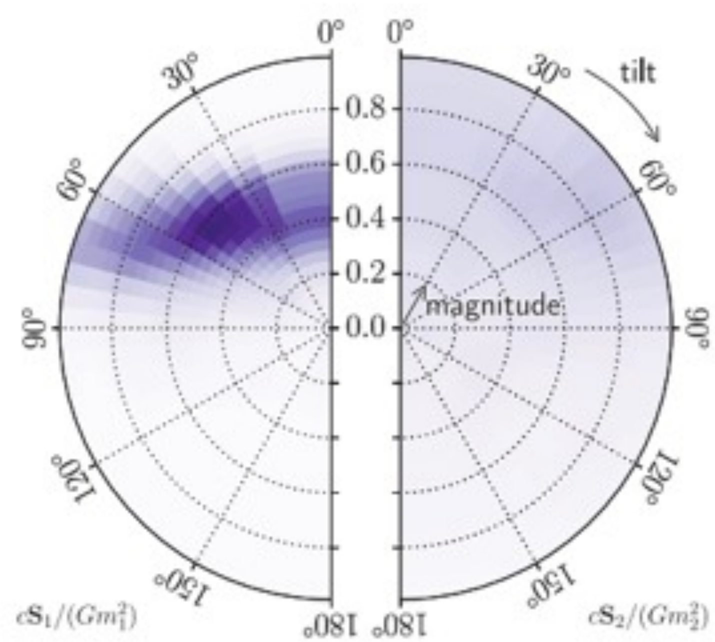


Parameter Degeneracies:
Primarily sensitive to the *chirp mass* — leaves **large degeneracies** along contours of chirp mass (**GW151226** approaching $m_2 < 3$ region)

$$\mathcal{M}_c = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}$$

$$\chi_{\mathrm{eff}} = \frac{m_1 s_{1,z} + m_2 s_{2,z}}{m_1 + m_2}$$



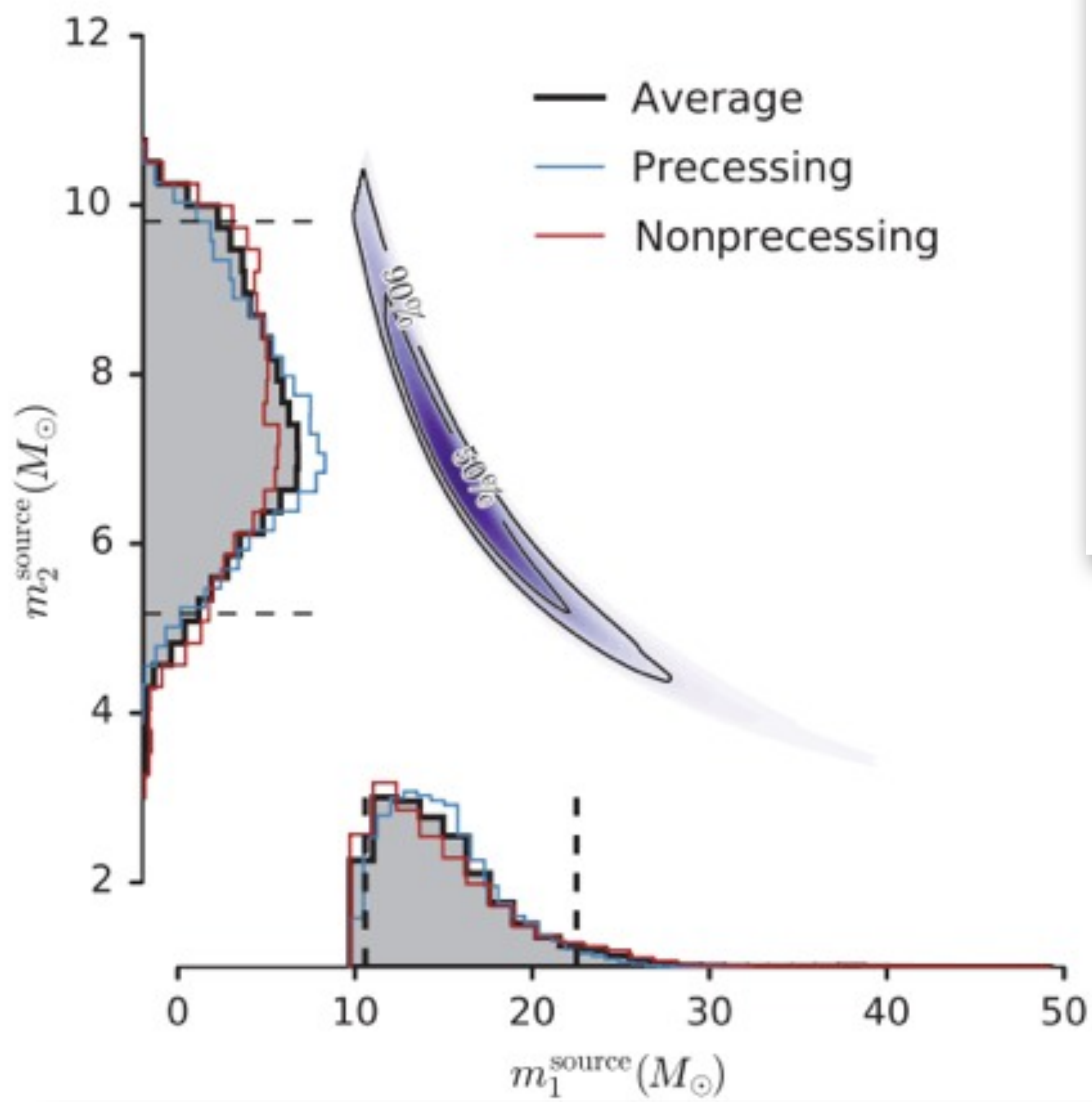


Frequency content (and thus "length in band" affected by both *effective spin* and *mass ratio* at same order in expansion of radiation amplitude/phase

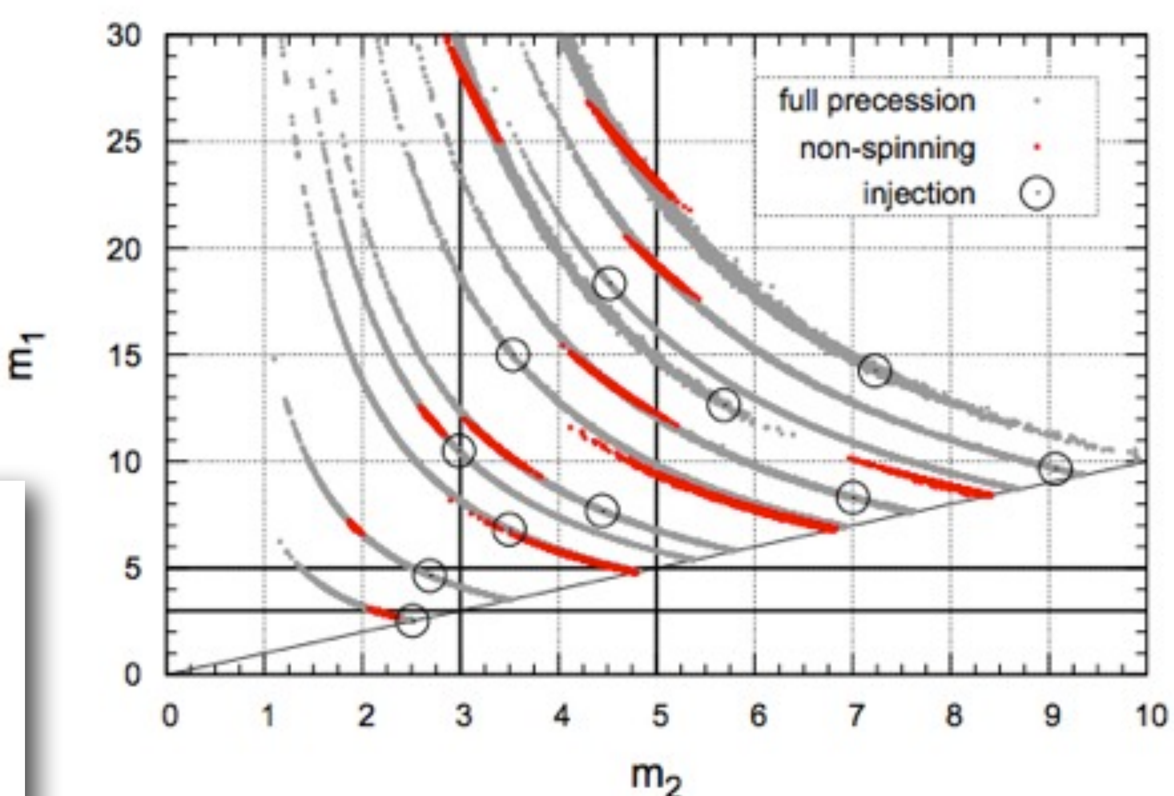
Phys. Rev. Lett. 116/241103

Systematics In Parameter Estimation

Phys. Rev. Lett. 116/241103



Systematic Errors: We don't know what the **right** answer is. Hit it with the best waveform models we have, covering a wide range of physical features and calibrations to numerical relativity. The answers end up being similar for common parameters — careful and exhaustive study of differences could give hints at underlying physical processes!



This is costly, especially in higher dimensions... need a good balance between full information and resources. Some approximations bias other physical measurements

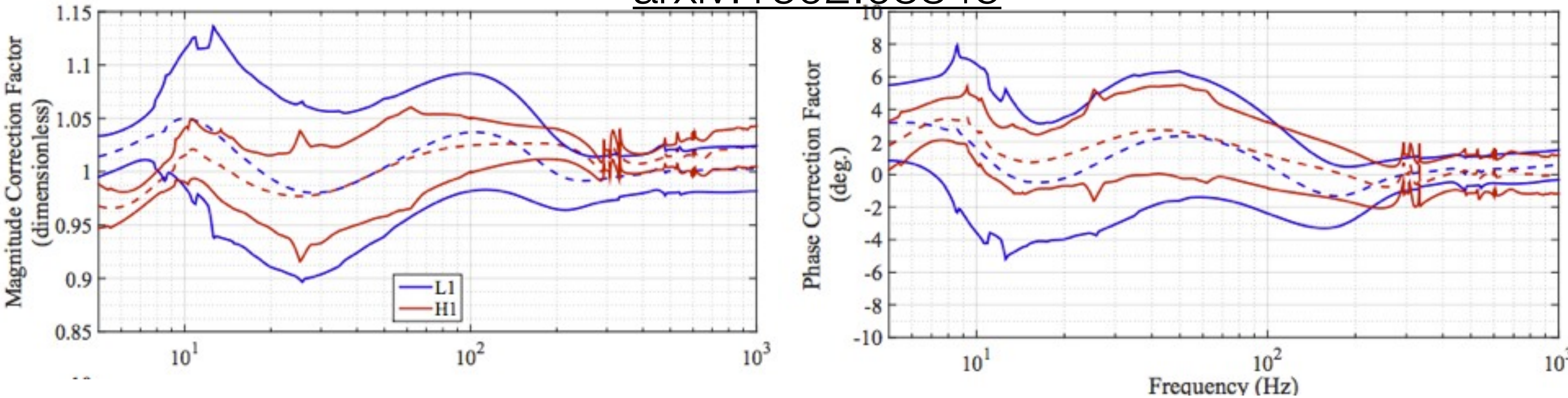
arxiv:1601.02661

Calibration Uncertainty

Problem: In reality, the strain measurement is derived from a differential phase between two (nominally) coherent laser beams. We model the instrument response at different frequencies to derive h from phase measurement. How do we deal with measurement and calibration error?

$$h(f) = A(f)e^{\imath\phi(f)} \to (A(f) + \delta A(f))e^{\imath(\phi(f)+\delta\phi(f))}$$

arxiv:1602.03845

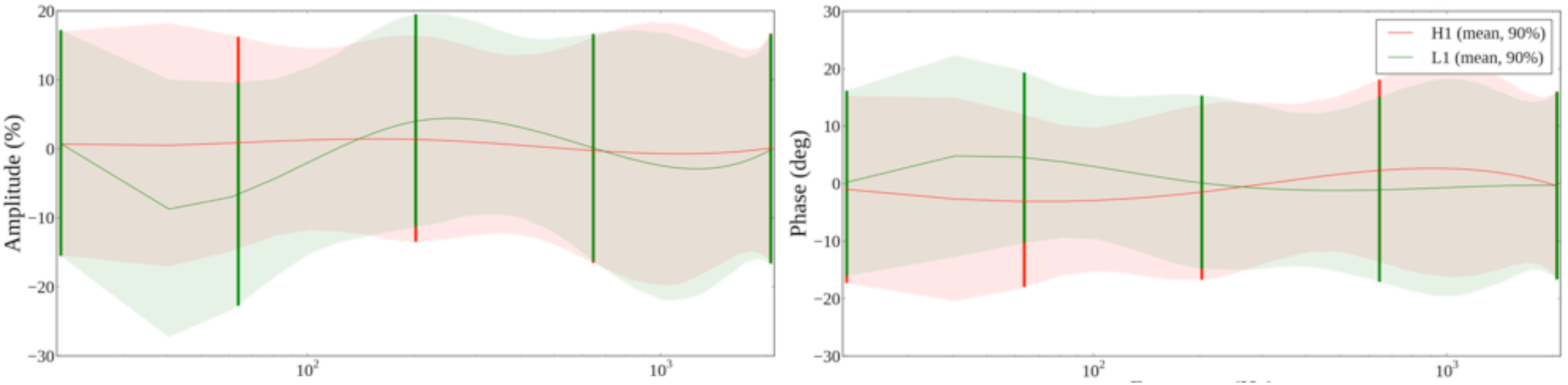


We can empirically measure the error: typically of order 5-10% in amplitude and few degrees in phase (very frequency dependent)

Calibration Uncertainty

Problem: In reality, the strain measurement is derived from a differential phase between two (nominally) coherent laser beams. We model the instrument response at different frequencies to derive h from phase measurement. How do we deal with measurement and calibration error?

$$h(f) = A(f)e^{\imath\phi(f)} \rightarrow \left(A(f) + \delta A(f)\right)e^{\imath(\phi(f)+\delta\phi(f))}$$



Model: Incorporate the amplitude/phase uncertainties into our Bayesian model as a set of parameters to estimate. The overall uncertainty is modeled a *spline fit* with control points in frequency space and *errors* attached to each point in *relative amplitude and phase* (simulated noise shown here)

Background Sample

Problem: Our noise is **not** Gaussian — it is contaminated with environmentally induced transients; many of which can not be safely excluded with data quality concerns. How do we model the background?

Model 1: Slide instrument data with respect to each other, breaking time-coincidence (and hence one of our signal model assumptions) — build up coincidence events from the slides into a distribution in ranking statistic (SNR)

Model 2: Build up a likelihood ratio ranking statistic from non-coincident event triggers and an analytical model of expected signal distributions

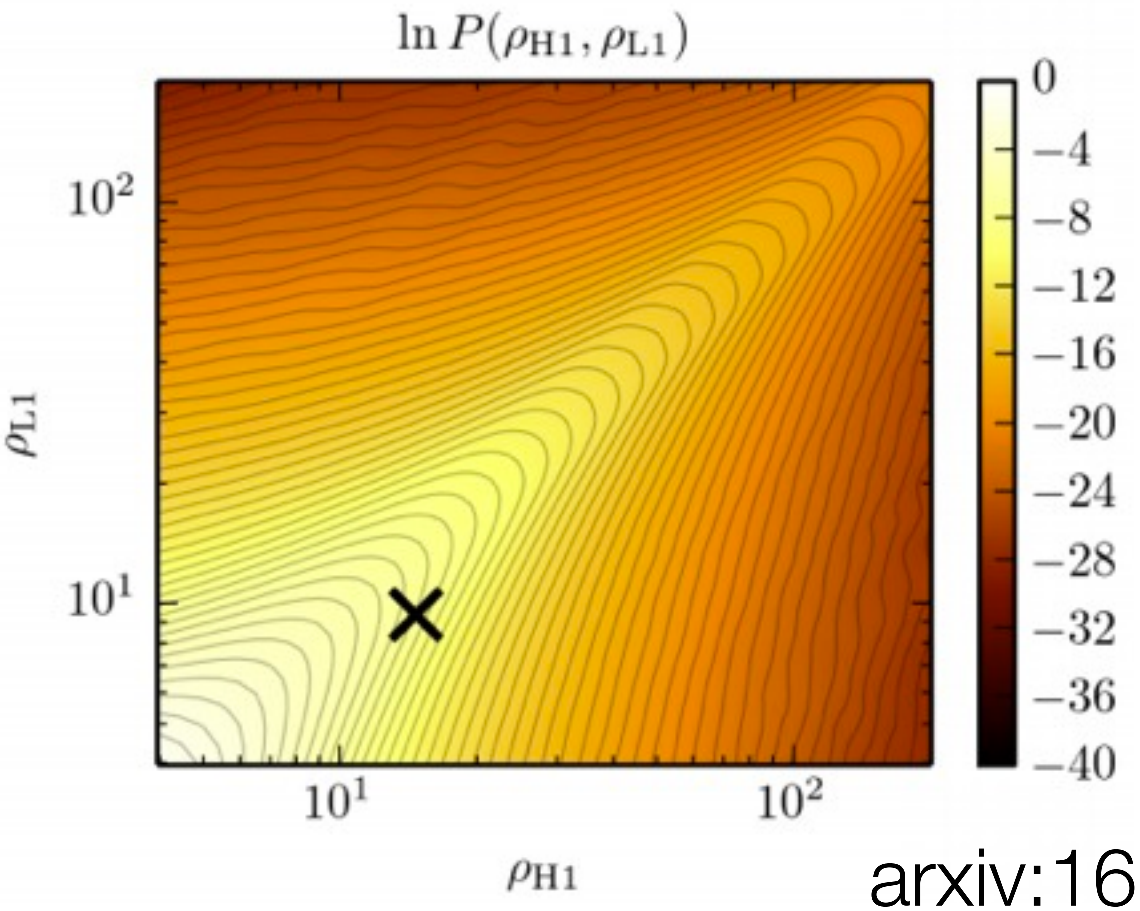
$$p(\rho) = \frac{\lambda(p)^n e^{-\lambda(\rho)}}{n!}$$

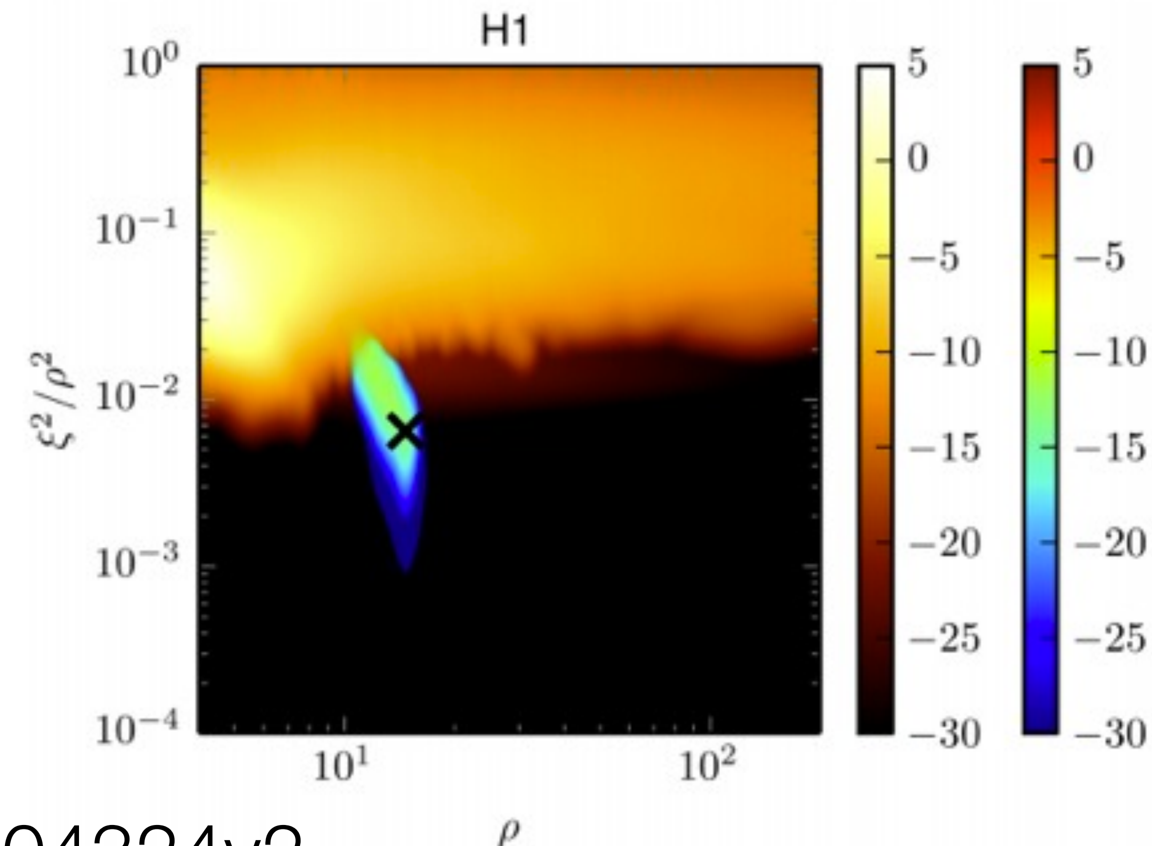
$$\mathcal{L} = \frac{p(\rho_H, \chi^2_H, \rho_L, \chi^2_L, \cdots | h)}{p(\rho_H, \chi^2_H, \rho_L, \chi^2_L, \cdots | n)}$$

Model 1: $\lambda(\rho) \sim R(\rho) \times T_{obs} / N_{slides}$ — the expected number at a given ranking statistic value is now measured from the background population

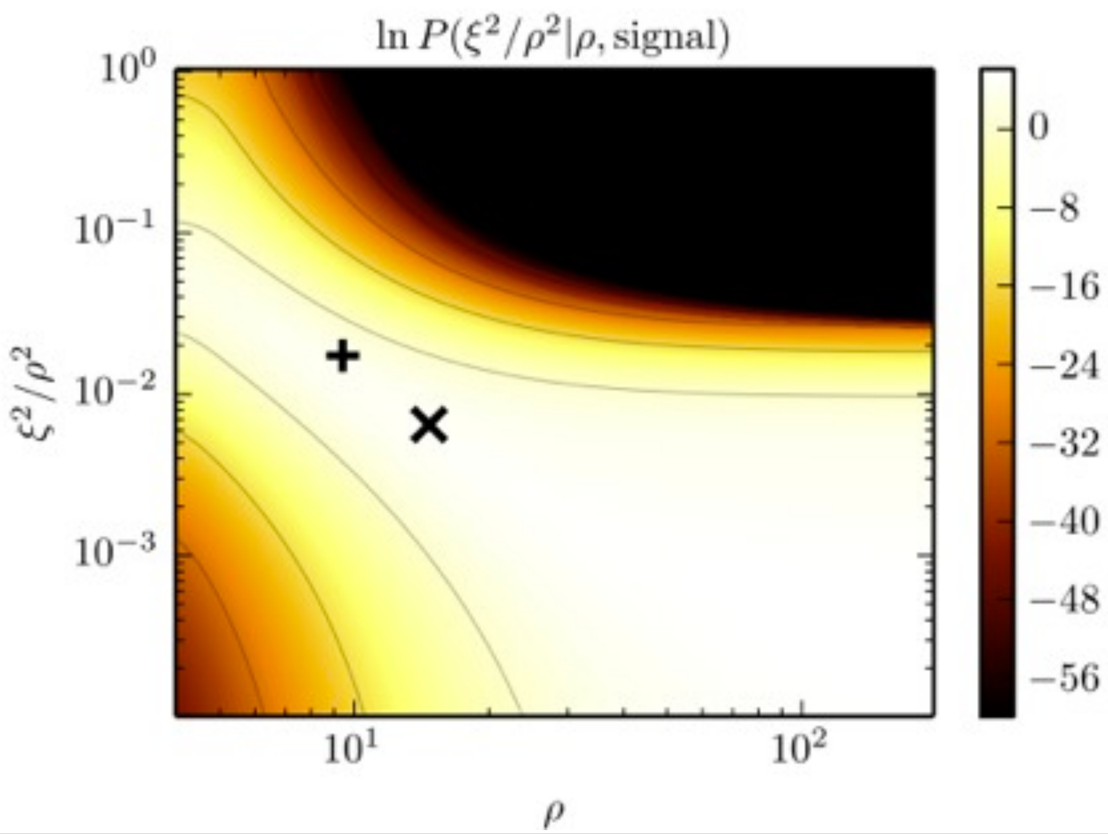
Model 2: Numerator is analytical and calculated almost directly from $P(\rho|h)$, but with the modeled expectation from multiple detectors. The denominator is factored into individual instruments and determined empirically

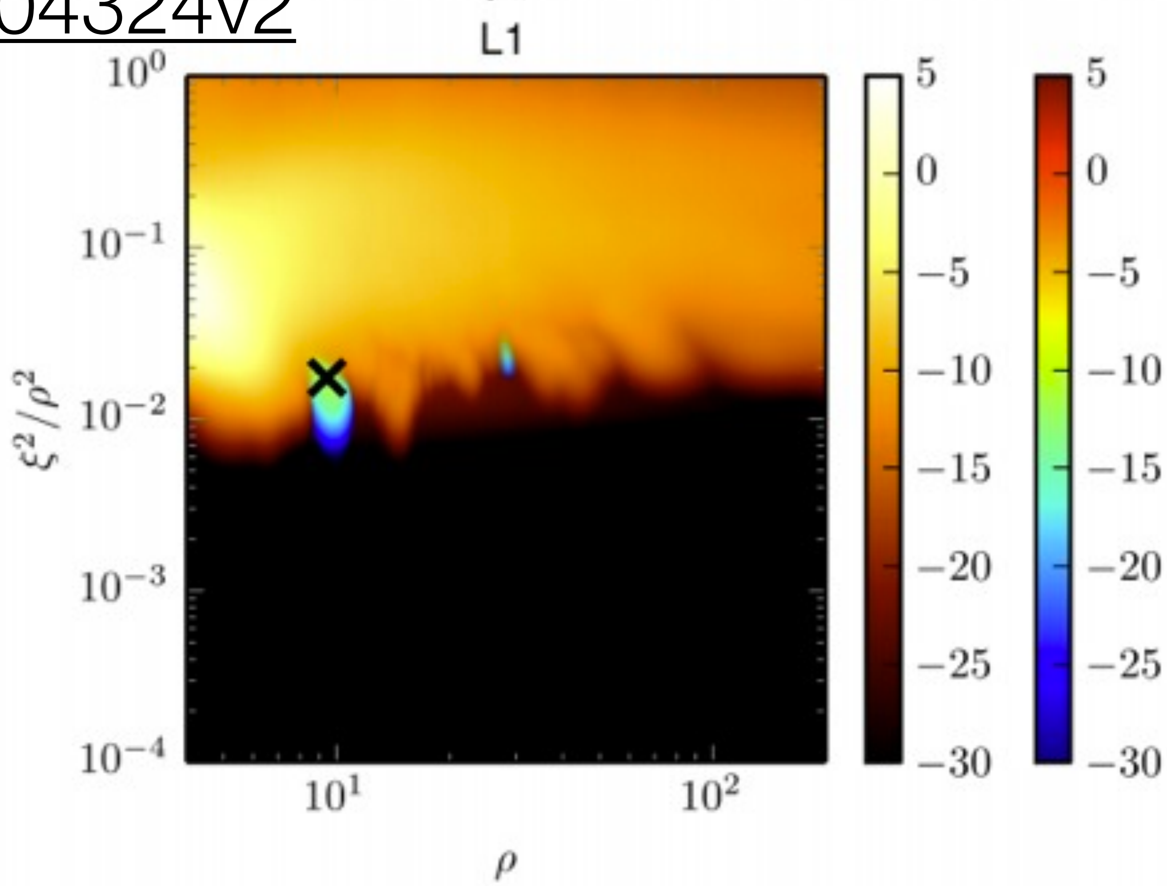
Likelihood Ratio Ranking Statistic





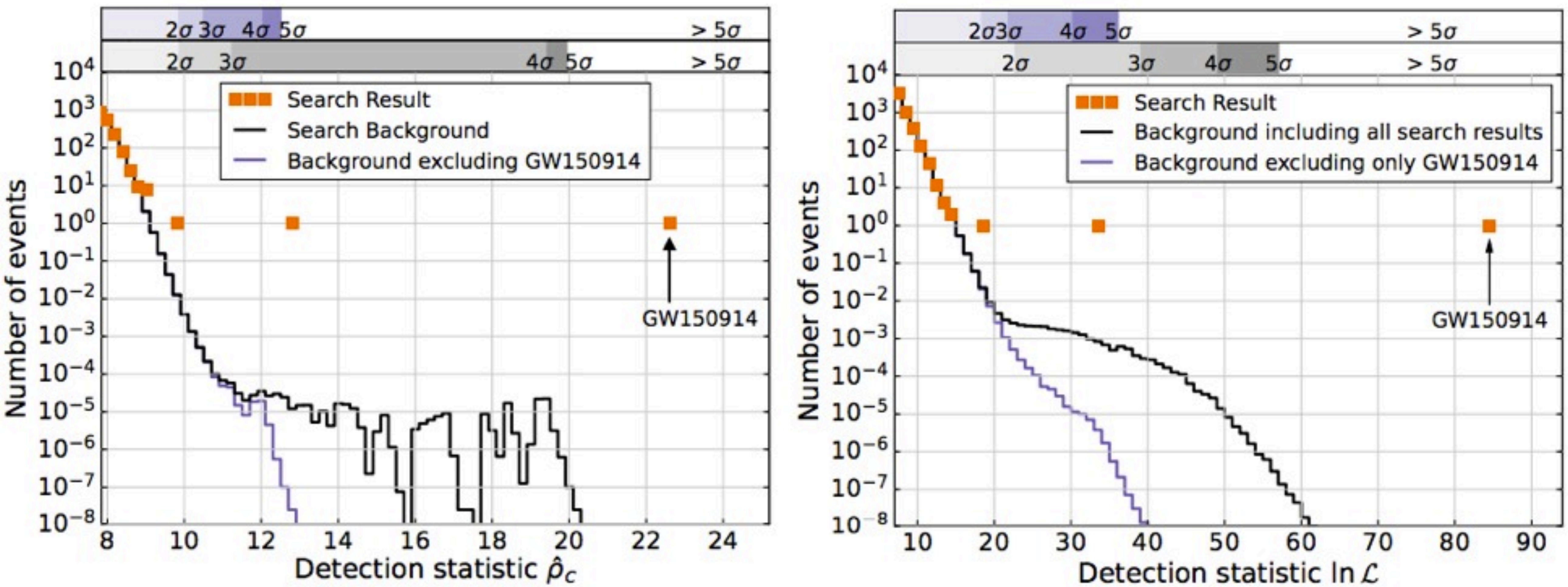
arxiv:1604.04324v2





Background Sample

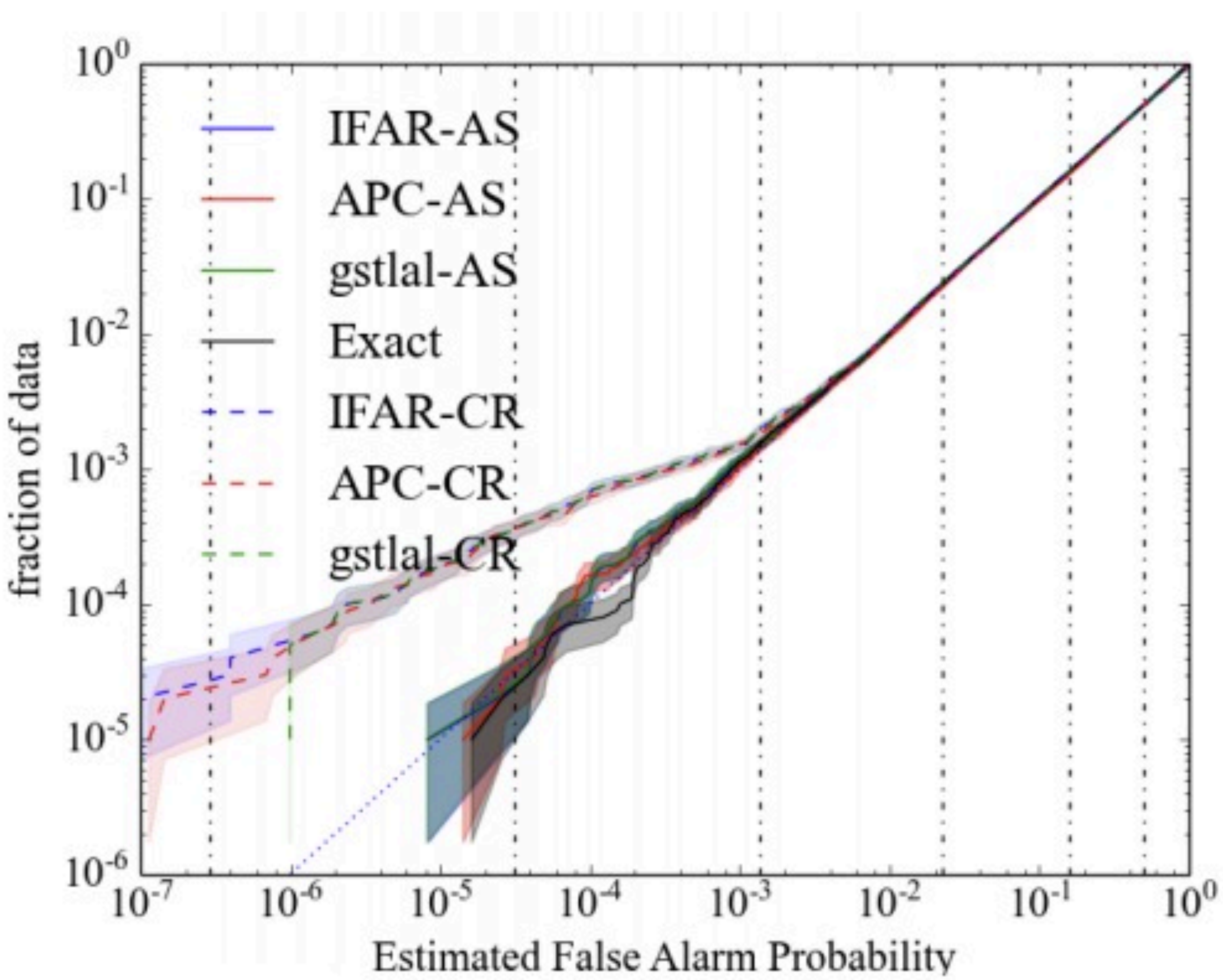
Problem: Our noise is **not** Gaussian — it is contaminated with environmentally induced transients; many of which can not be safely excluded with data quality concerns. How do we model the background?



Answer: In both models, our background is estimated by constructing an estimate of the rate of coincident triggers from the "no gravitational wave present" hypothesis set, but...

Event Significance

Problem: One cannot shield against gravitational waves (with current budgets). However, in order to establish significance of a given event, However, how does one contend with background contamination from the signal?



Solid lines represent various methods (giving mostly similar results) without signal removal. **Dashed lines** do remove the signal before calculating a false alarm probability. **Shaded regions** are uncertainty equated with Poisson process error bars

Answer: We don't. A controlled study shows that methods which remove the signal from its own background end up biasing detection confidence (e.g. p-values)

arxiv:1601.00130

Inferred Rates / Probability of Astrophysical Origin

$$\mathscr{L} = \left\{ \prod_i \Lambda_{\rm bg} p_{\rm bg}(x_i) + \Lambda_{\rm fg} p_{\rm fg}(x_i) \right\} \exp(-\Lambda_{\rm bg} - \Lambda_{\rm fg})$$

Likelihood of obtaining ensemble of ranking statistics $\boldsymbol{x_i}$ with two categories of events:
background (terrestrial) and foreground (astrophysical)
$\Lambda_{fg,bg}$ ~ expected counts from each category
p_{fg}, p_{bg} - modeled or measured, for astrophysical distribution of binaries $p_{fg} \sim \rho^{-4}$
Methods using LR ranking can divide out p_{bg} and use likelihood statistic directly

$$p(\Lambda_{\rm bg}, \Lambda_{\rm fg}) = \frac{1}{\sqrt{\Lambda_{\rm bg}\Lambda_{\rm fg}}}$$

Obtain posterior on \Lambda which scales with the rate by the sensitive space-time volume by marginalization over the x_i, applying a Jeffrey's prior on the rates

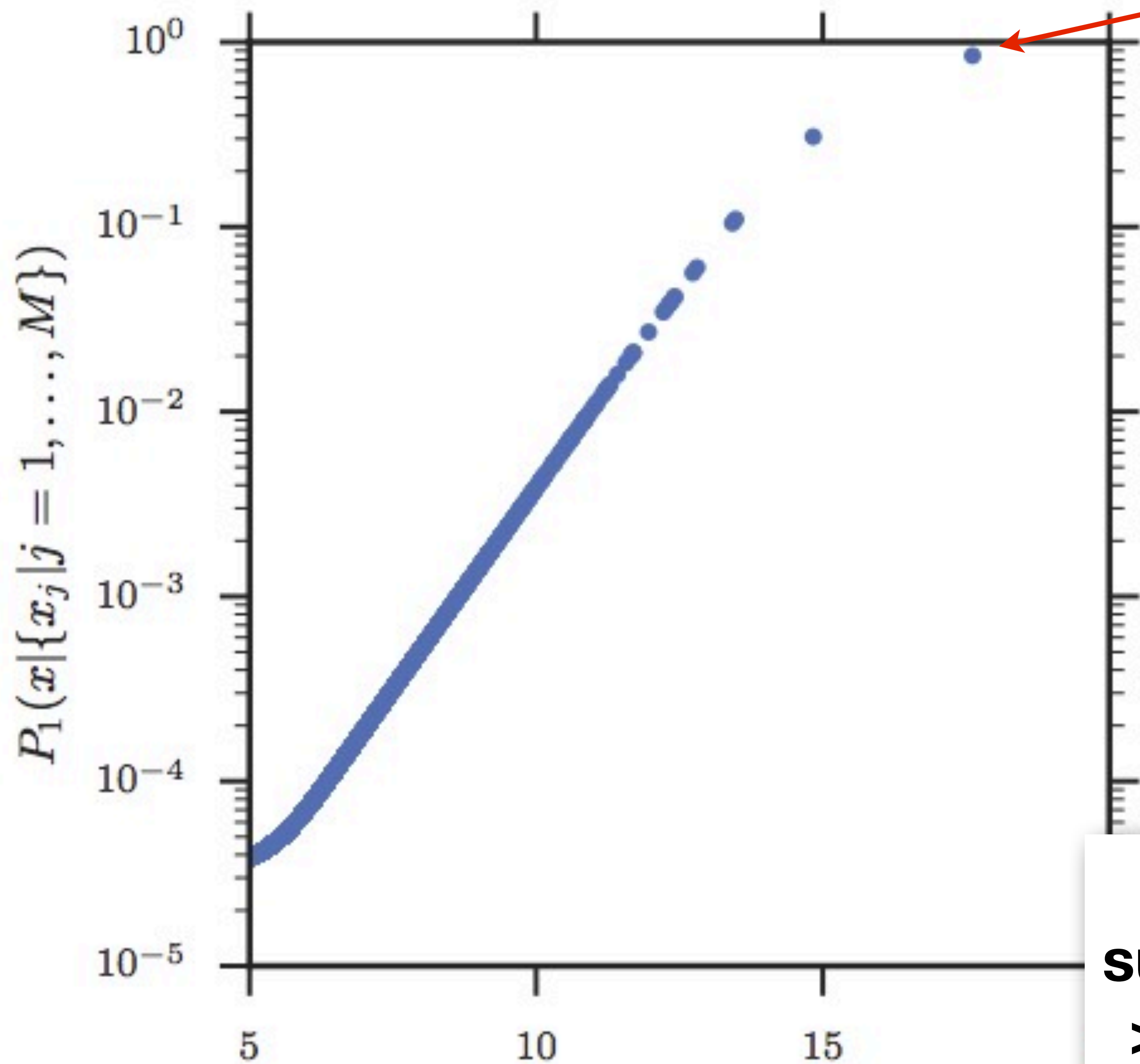
arxiv:1302.5341

Inferred Rates / Probability of Astrophysical Origin

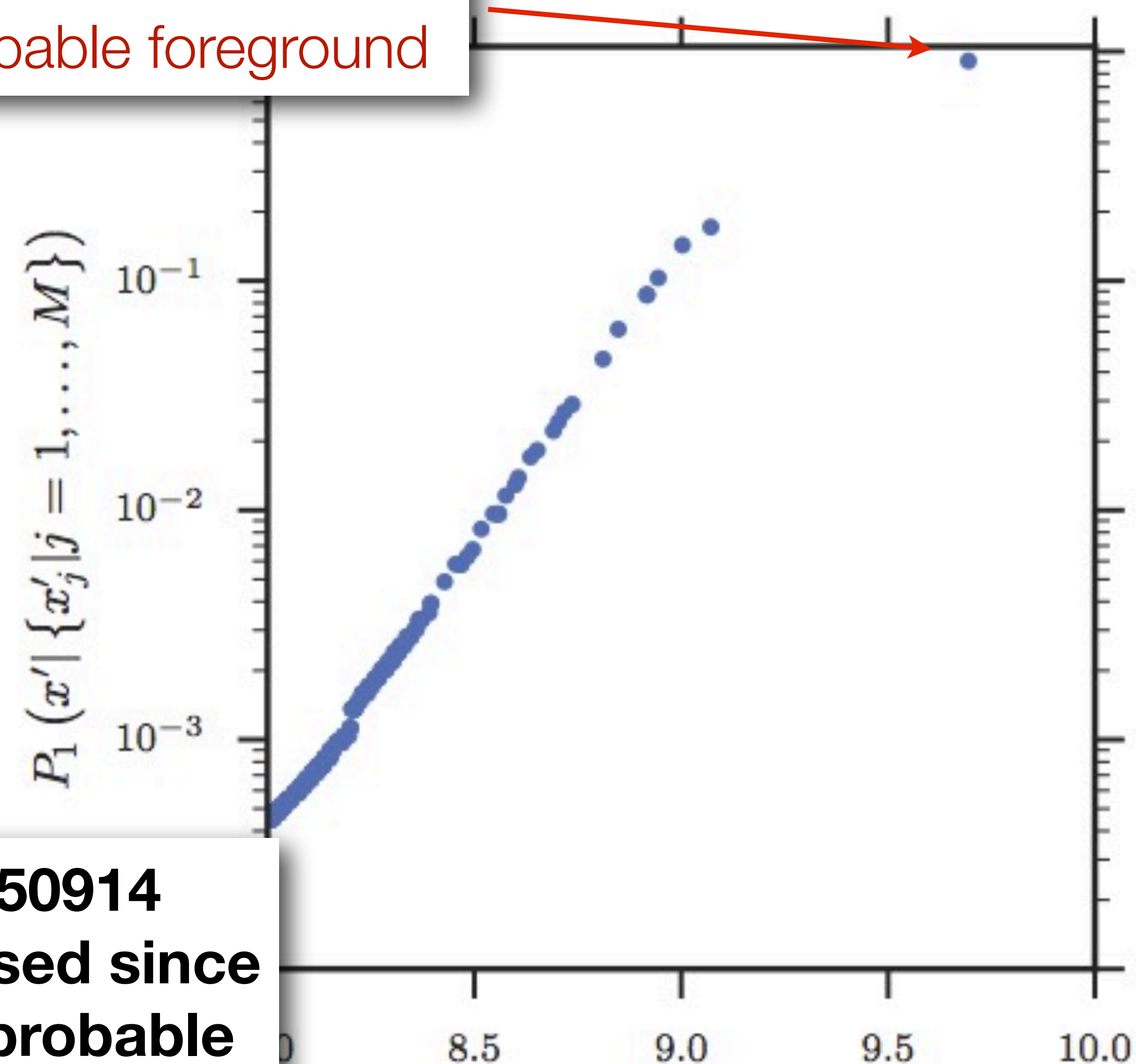
Obtain probability of astrophysical origin by marginalizing against the counts

$$p_{\rm astro}(x|x_i) = \int d\Lambda_{\rm fg} d\Lambda_{\rm bg} \frac{\Lambda_{\rm fg} p_{\rm fg}(x)}{\Lambda_{\rm fg} p_{\rm fg}(x) + \Lambda_{\rm bg} p_{\rm bg}(x)} p(\Lambda_{\rm fg}, \Lambda_{\rm bg}|x_i)$$

LVT151012 ~ 87% probable foreground



Model 2 ranking stat.



Model 1 ranking stat.

GW150914 suppressed since > 99% probable and far to the right

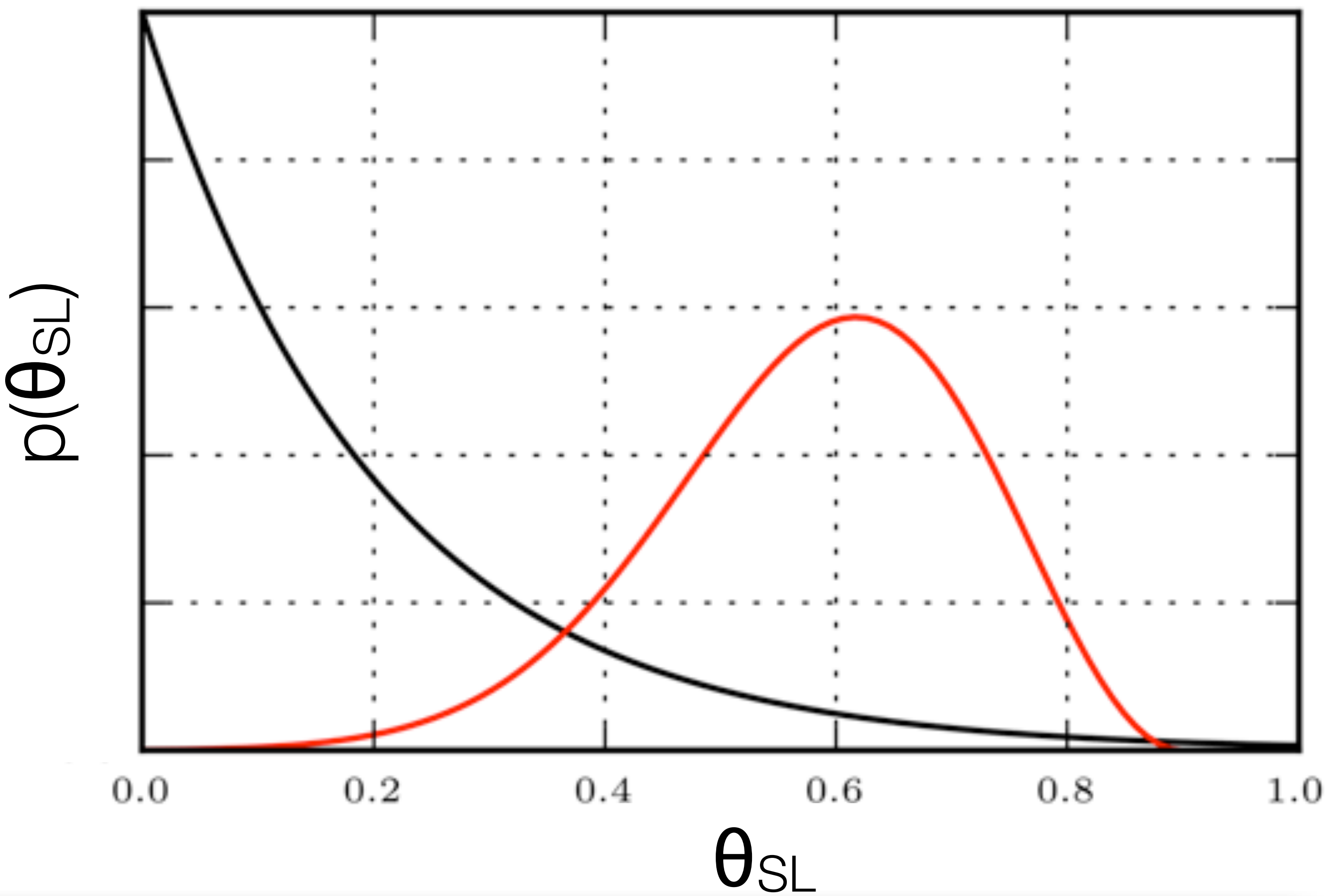
Towards Hierarchical Modeling

Problem: How do we measure populations with different characteristics? For example: we expect co-evolved binaries to have aligned BH spins, and cluster capture/dynamics to have random spins. Can we integrate this information and model it with our data?

$$\theta = \{\mathcal{M}_c, \theta_{SL}\}$$

$$\Lambda_{\mathrm{fg}} \rightarrow \Lambda_{\mathrm{fg}}(\theta)$$

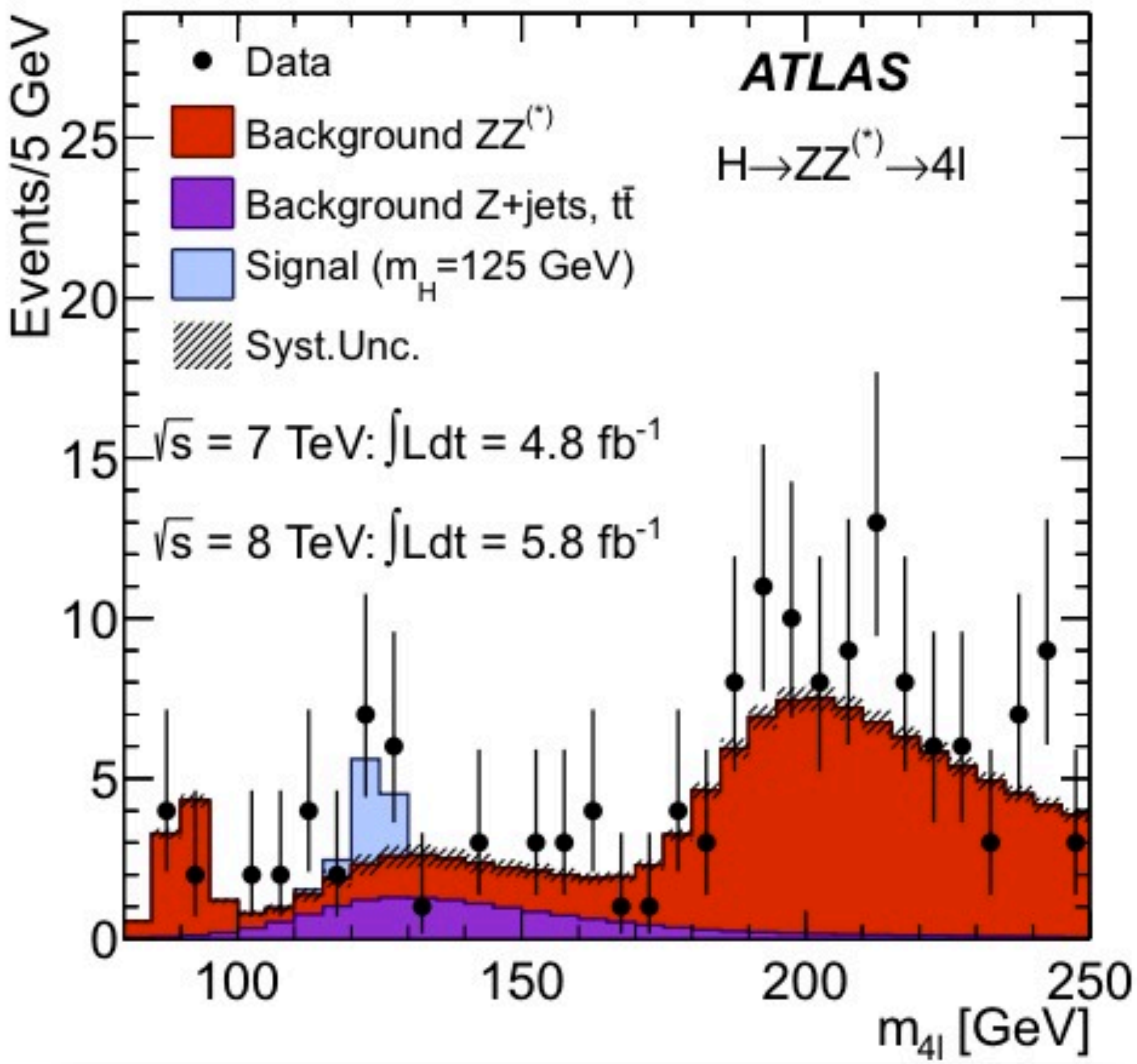
$$p(\Lambda_{\mathrm{fg}}, \Lambda_{\mathrm{bg}}) \rightarrow p(\Lambda_{\mathrm{fg}}, \Lambda_{\mathrm{bg}}, \theta)$$

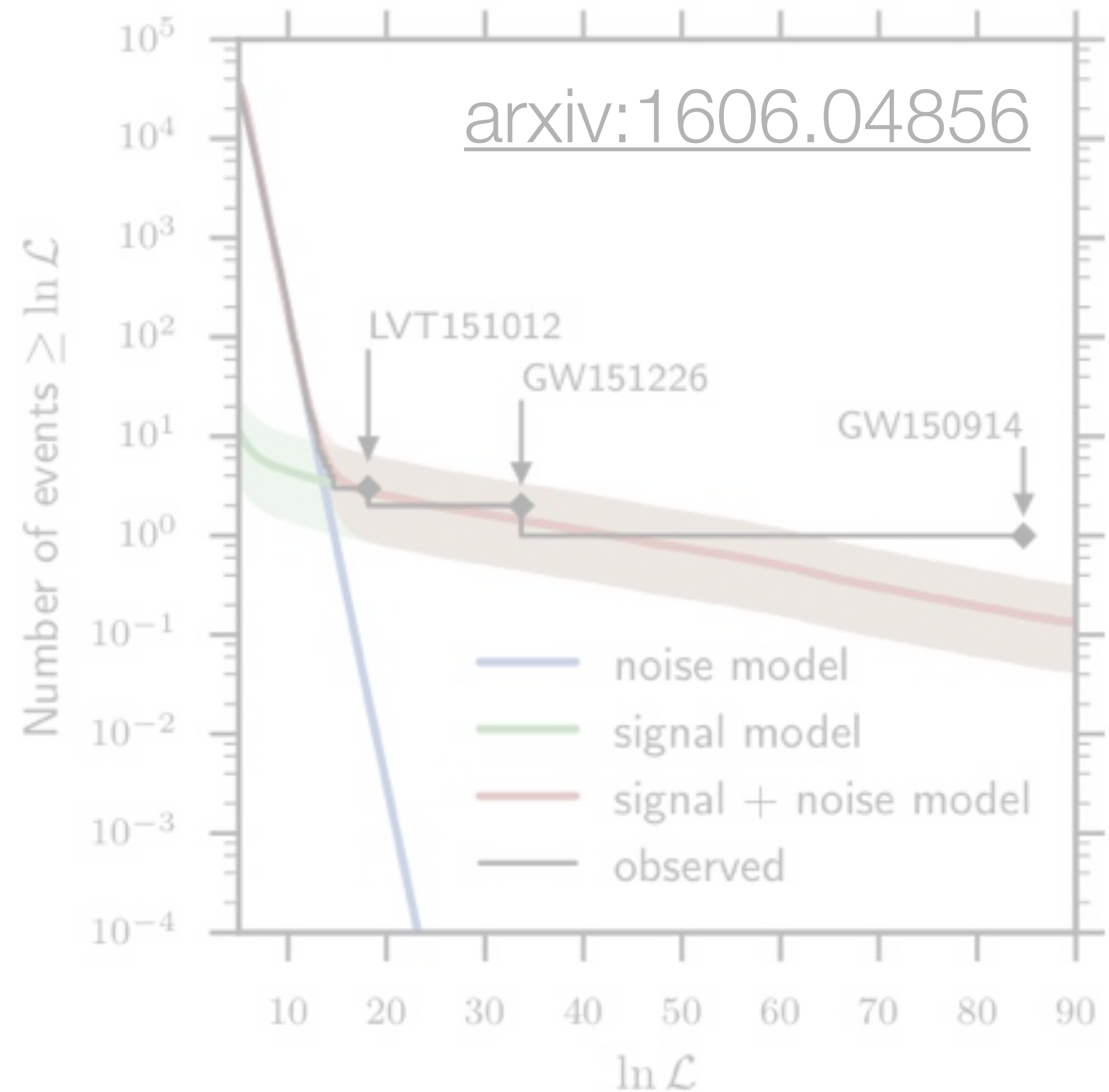


Answer: Integrate a set of parameters (θ) into our source classification and counts. Individual parameters like spin and mass distributions are folded naturally into the expected ranking distributions and we form posteriors from the ensemble over those parameters

BBH Detection

Phys. Lett. B (716) 1
Phys. Lett. B (716) 1



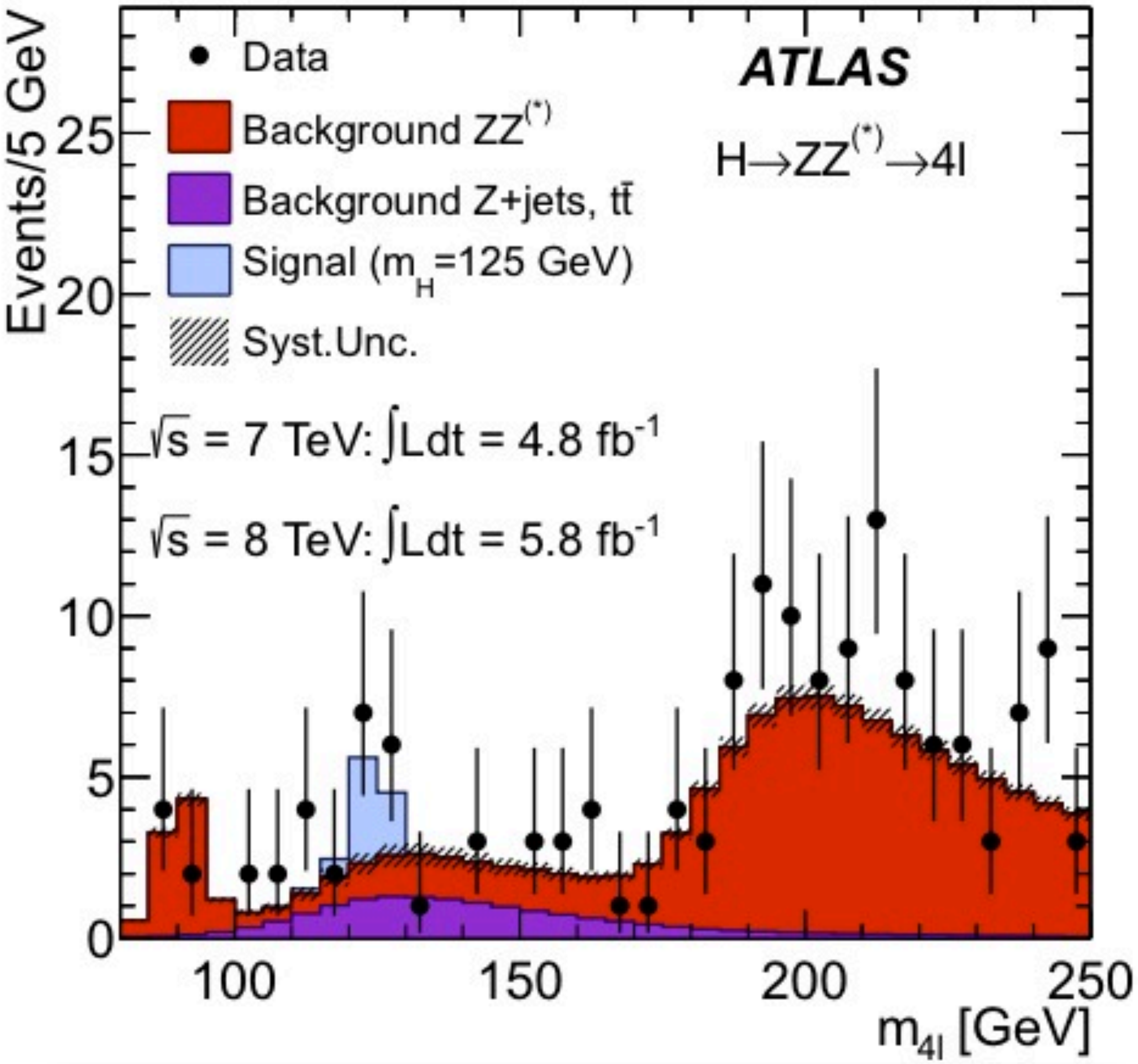


Signal and Background (Higgs):
For a given decay channel (4 lepton), this shows the background levels and expected Higgs signal decay rates along with the data collected — clear statistical excess ~125 MeV

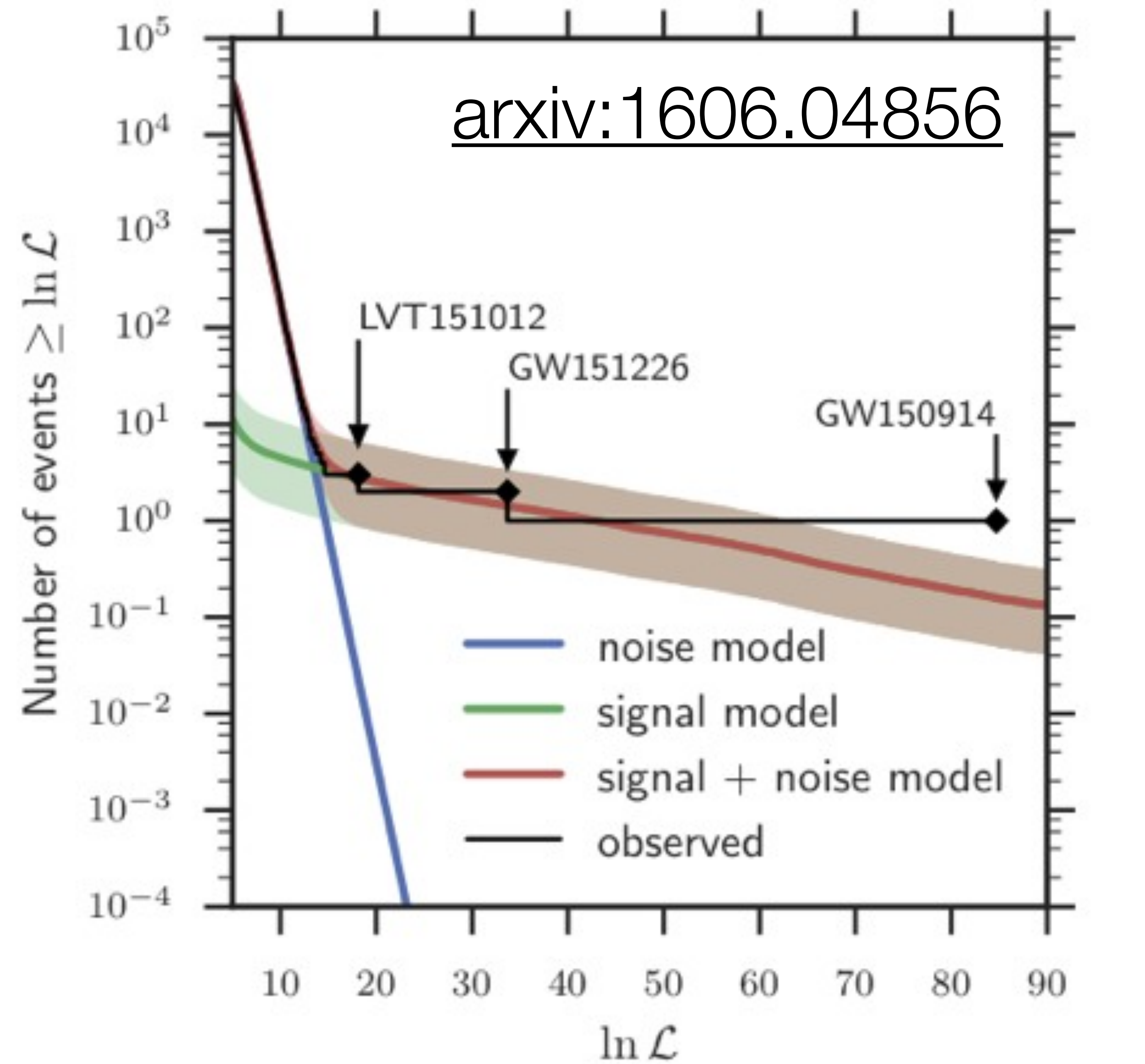
Signal and Background (GW):
Different parameterization, using a likelihood ranking statistic modeling background with the expected volumetric (ρ^{-4}) distribution superimposed

BBH Detection

Phys. Lett. B (716) 1
Phys. Lett. B (716) 1



arxiv:1606.04856

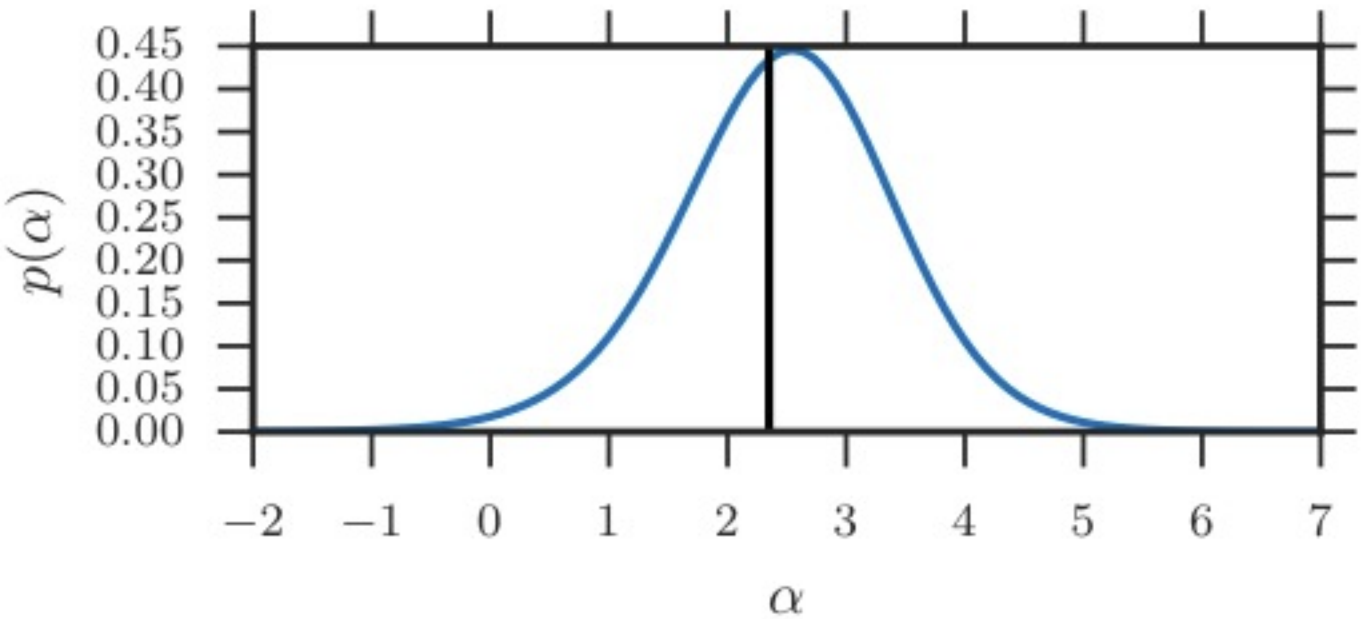


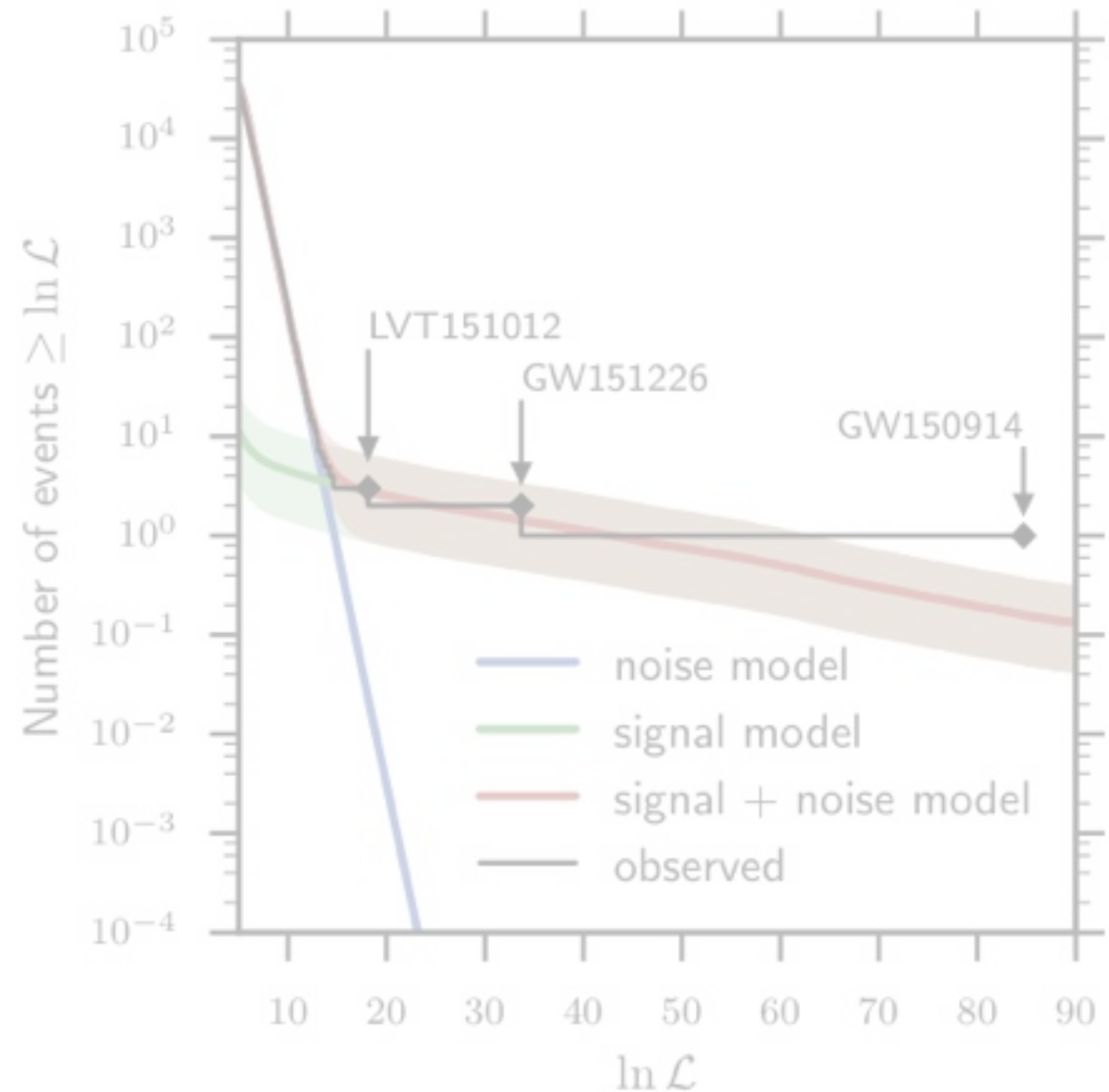
Signal and Background (Higgs):
For a given decay channel (4 lepton), this shows the background levels and expected Higgs signal decay rates along with the data collected — clear statistical excess ~125 MeV

Signal and Background (GW):
Different parameterization, using a likelihood ranking statistic modeling background with the expected volumetric (ρ^{-4}) distribution superimposed

BBH Detection

arxiv:1606.04856



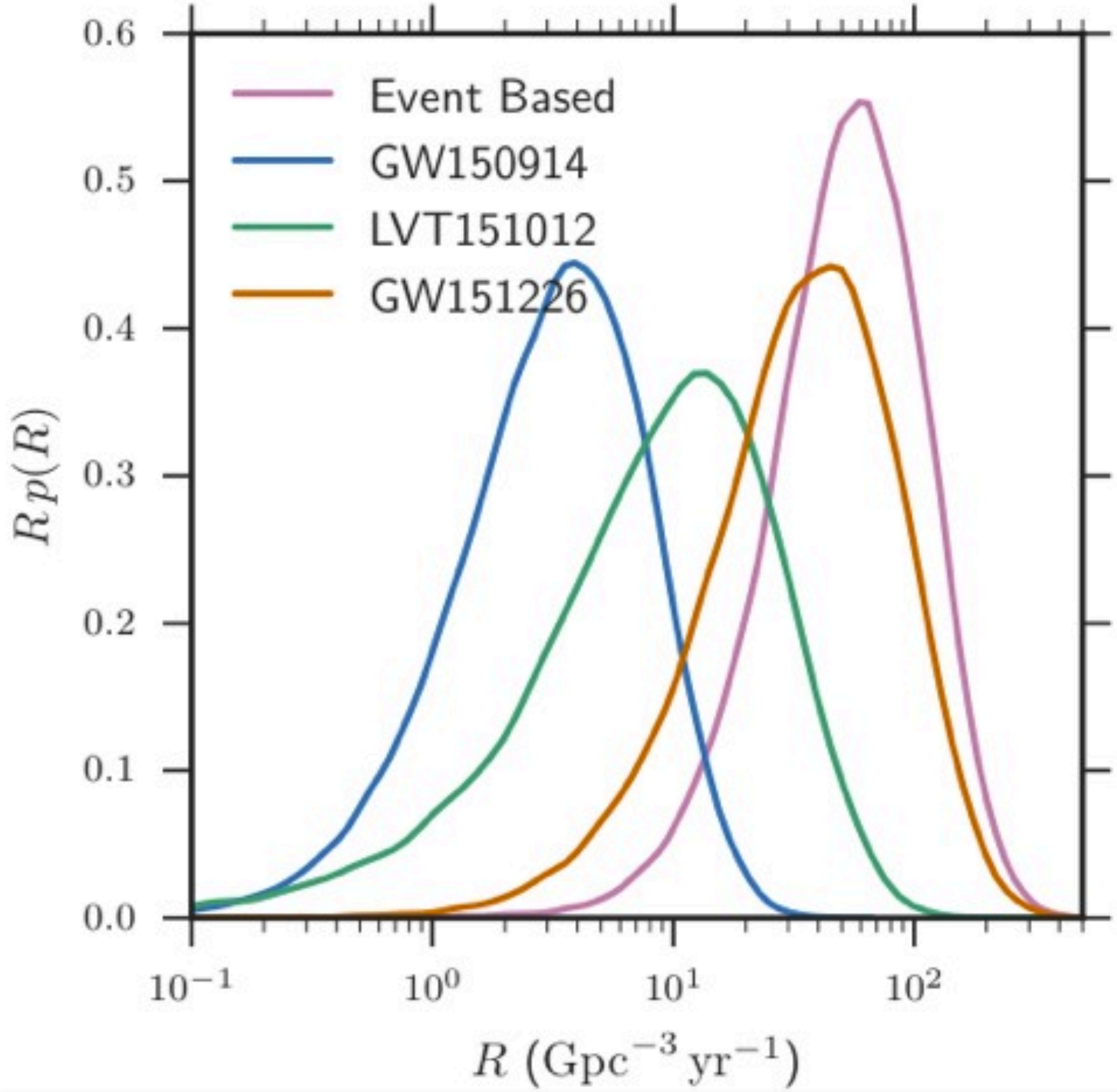


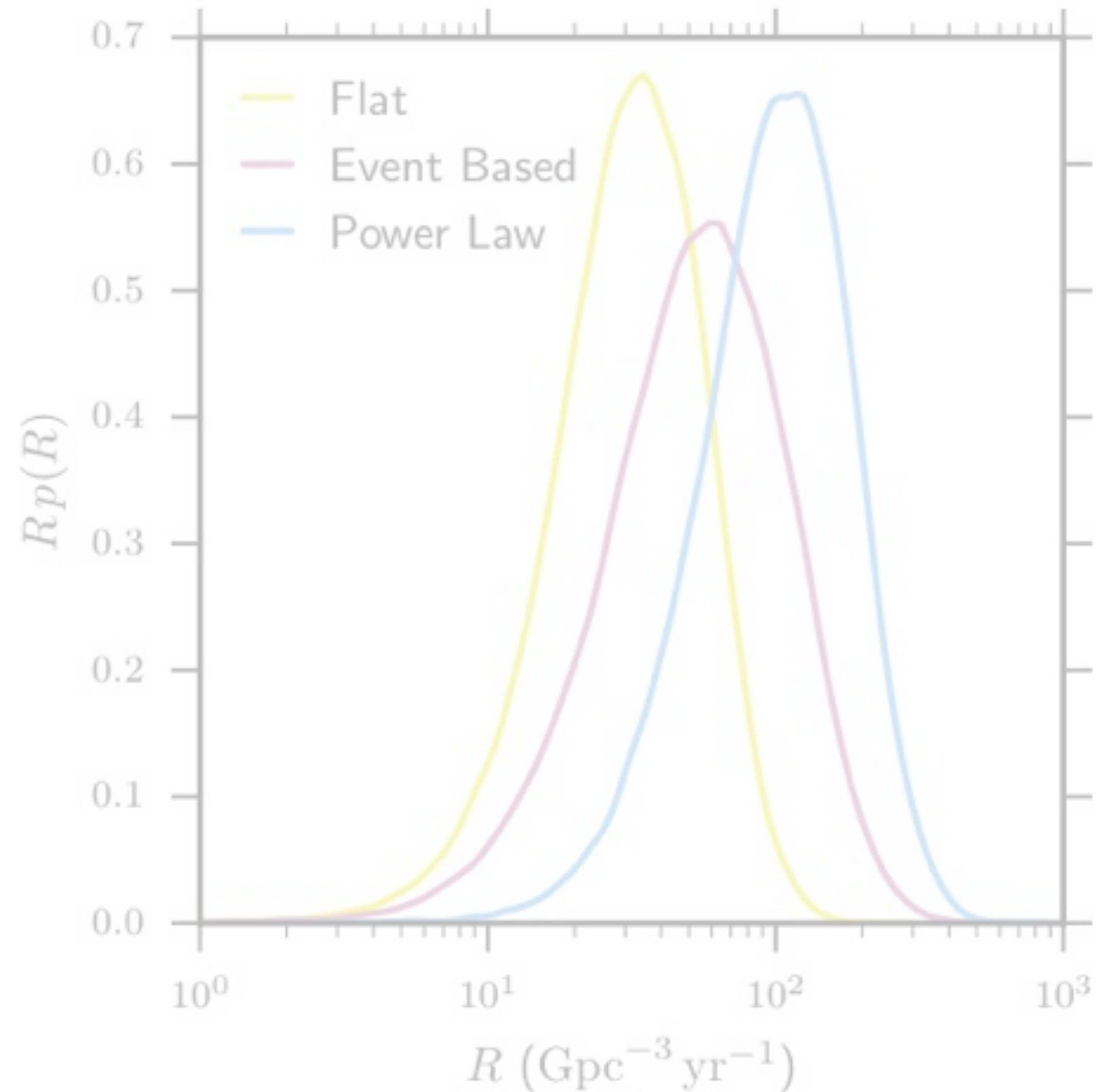
Basic Application of Hierarchical Modeling:
Posterior distribution for exponent of m_1 inferred from three astrophysically distinguished events — note peak very close to α = 2.35 (black vertical line)

Signal and Background (GW):
Different parameterization, using a likelihood ranking statistic modeling background with the expected volumetric (ρ^{-4}) distribution superimposed

BBH Event Rates

arxiv:1606.04856



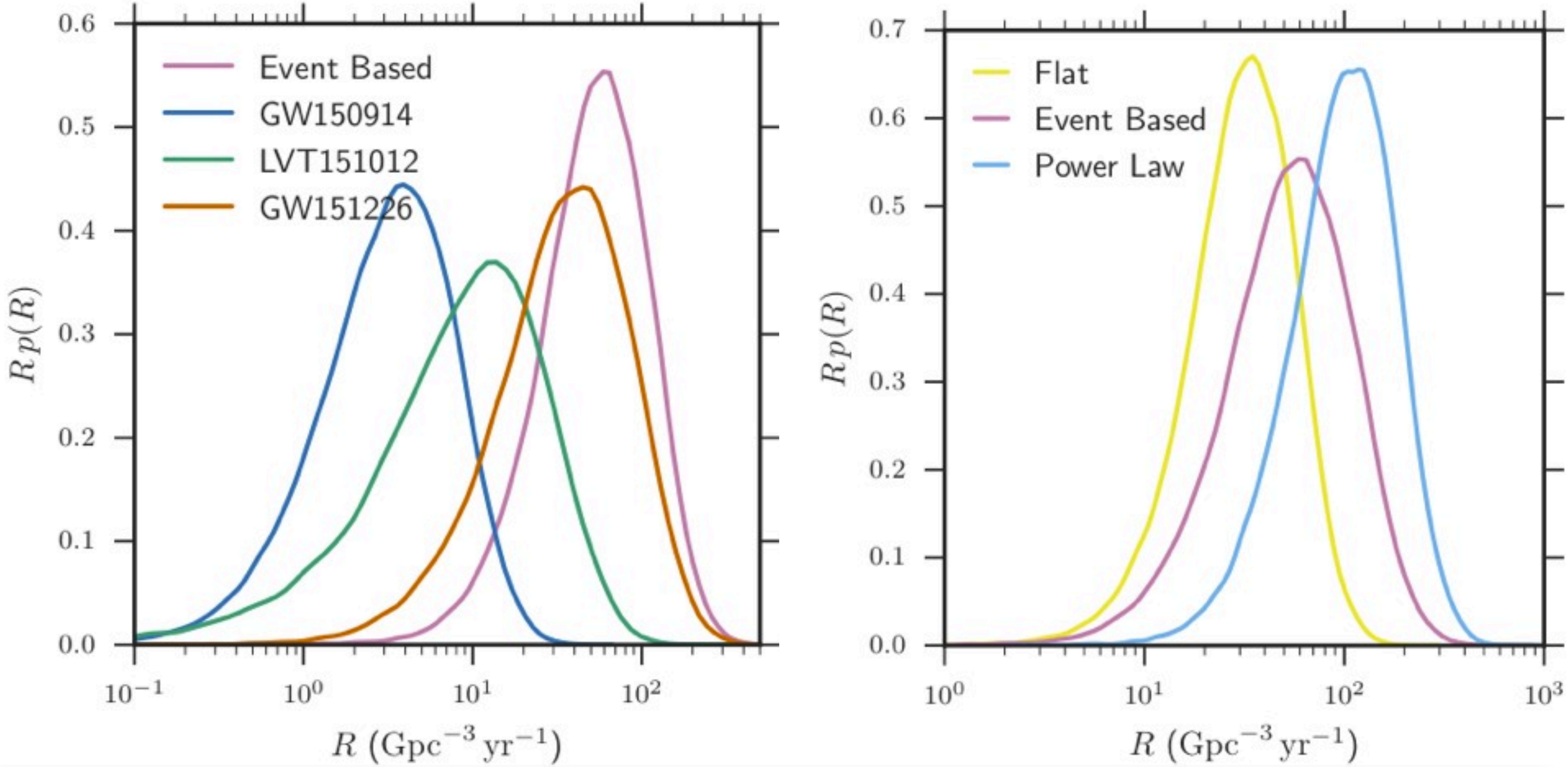


Dealing with Multiple Event Categories:

Being unsure of the intrinsic source populations and origins, we calculate the event rates for all three events and take the union to derive the overall event rate of BBH coalescence. Also test distributions of events according to uniform in the logarithm of component mass and according to the stellar initial mass function: $\mathbf{p(m_1) \propto m_1^{2.35}}$

BBH Event Rates

arxiv:1606.04856



Dealing with Multiple Event Categories:

Being unsure of the intrinsic source populations and origins, we calculate the event rates for all three events and take the union to derive the overall event rate of BBH coalescence. Also test distributions of events according to uniform in the logarithm of component mass and according to the stellar initial mass function: $\mathbf{p(m_1) \propto m_1^{2.35}}$

High Energy Neutrino Joint Search

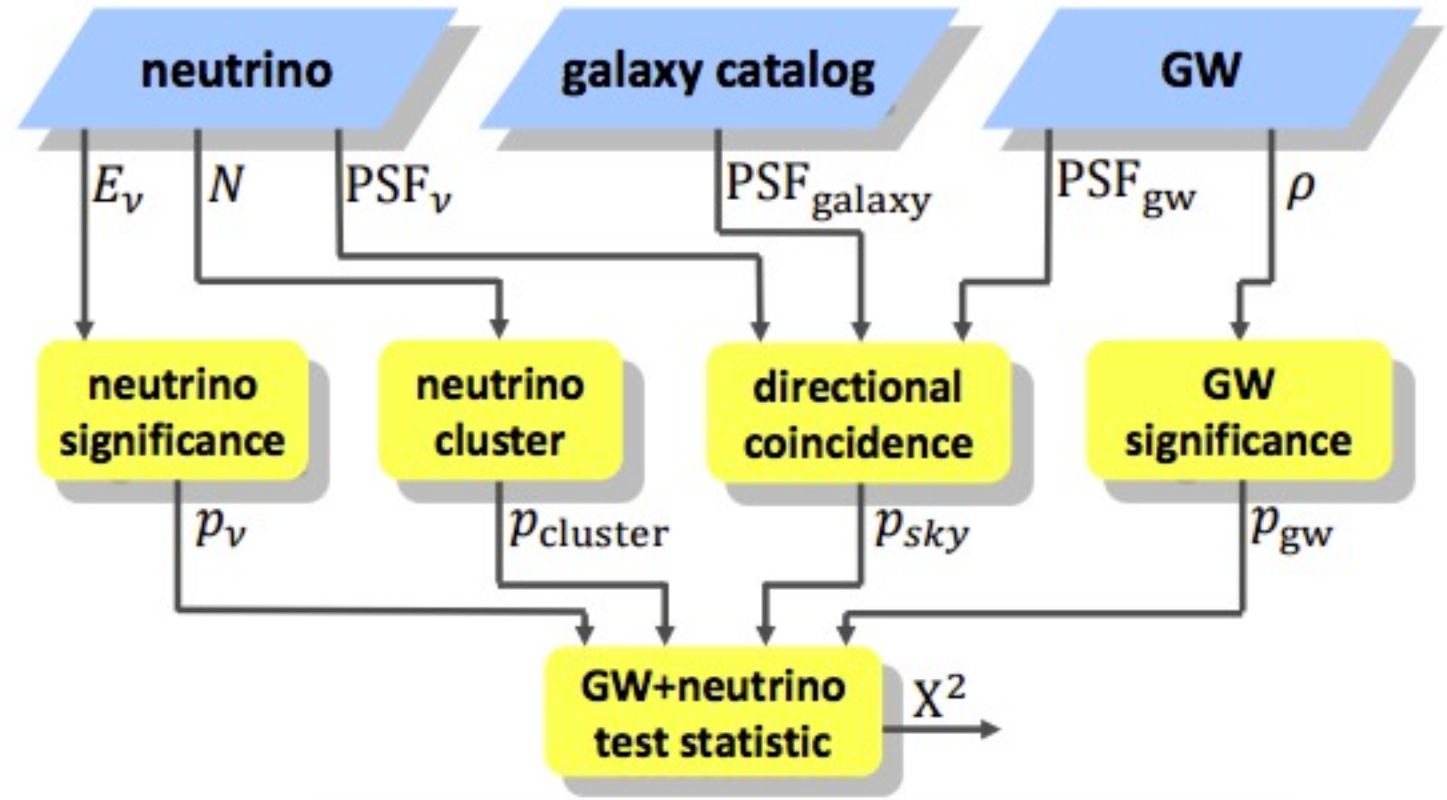
$$X_i^2 = -2\ln\left(p_{\mathrm{sky},i}\,p_{\mathrm{gw},i}\,p_{\mathrm{clus},i}\,p_{\nu,i}\right)$$

$$p_{\nu,i} = P(E_\nu^{\mathrm{BG}} \geq E_{\nu,i})$$

$$p_{\mathrm{clus}} \sim \mathrm{Pois}(k, f_{\nu,\mathrm{BG}} T_{\mathrm{wind}})$$

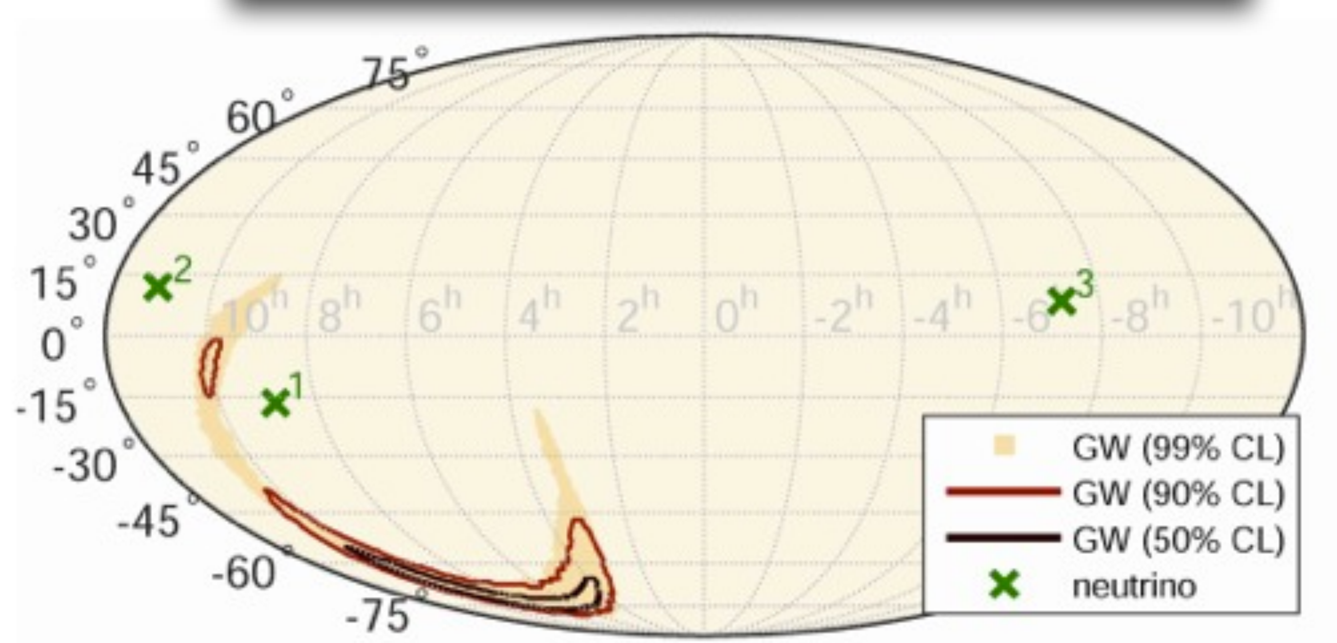
$$p_{\mathrm{gw},i} \sim \mathrm{Pois}(0, \lambda(\rho_i))$$

$$L_{\mathrm{sky}} \sim \int d\Omega\, p_{\mathrm{GW,gal}}(\alpha,\delta) \prod_{\nu_j} p_{\nu_j}(\alpha,\delta)$$



Multimessenger Searches:
Test statistic $\boldsymbol{X_i^2}$ (derived from Fisher's method) includes temporal (Poissonian) and sky coincidence with GW information and also folds in p-values derived from neutrino energy and probability of obtaining N > 1 neutrino

Sky coincidence with GW150914

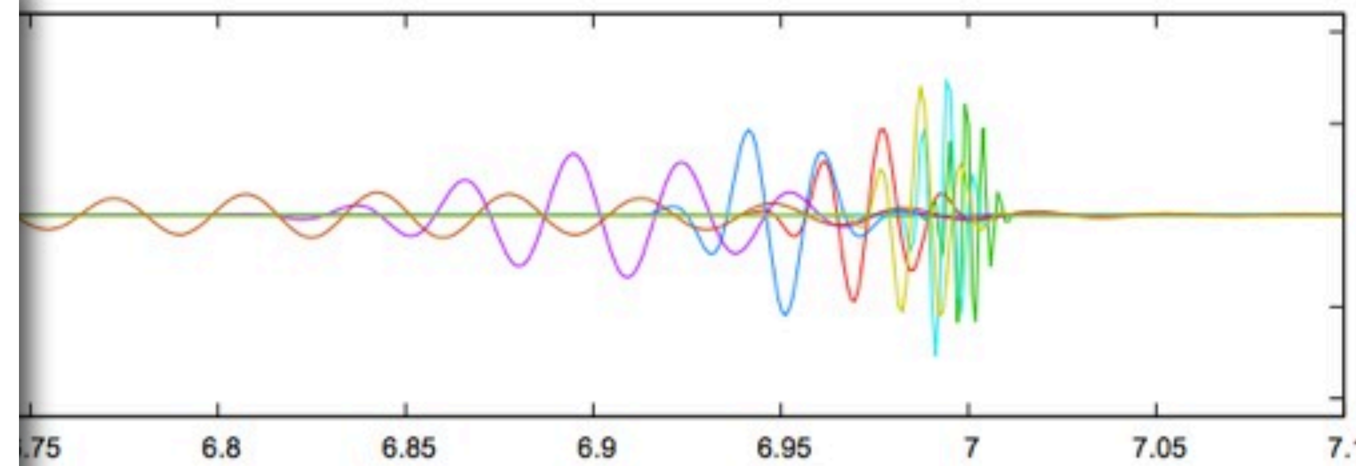


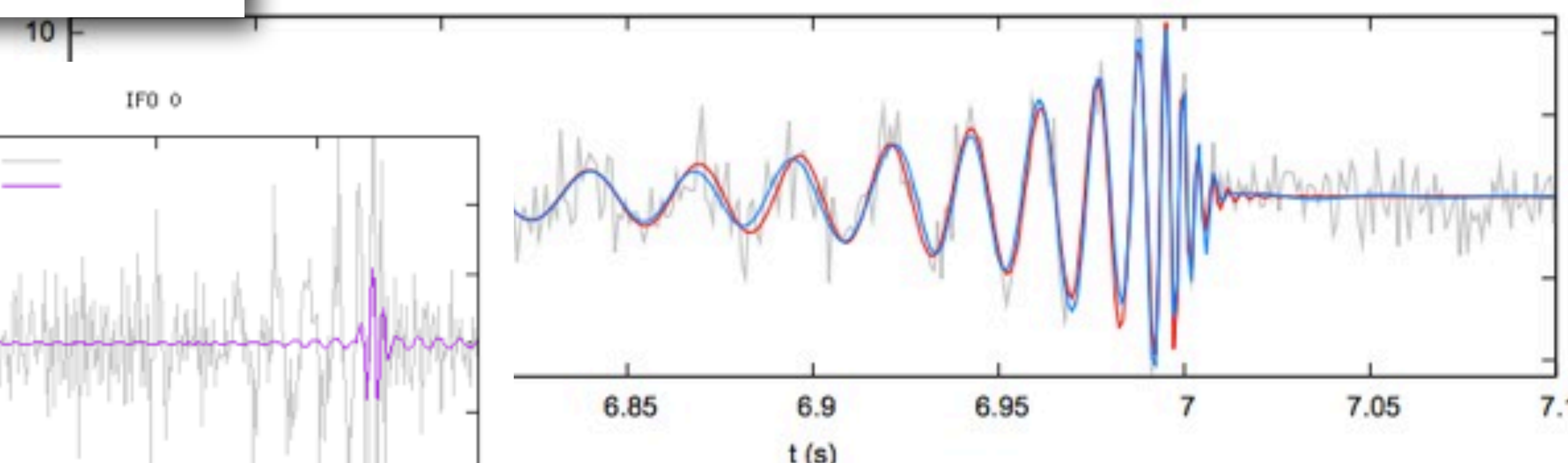
arxiv:1407.1042

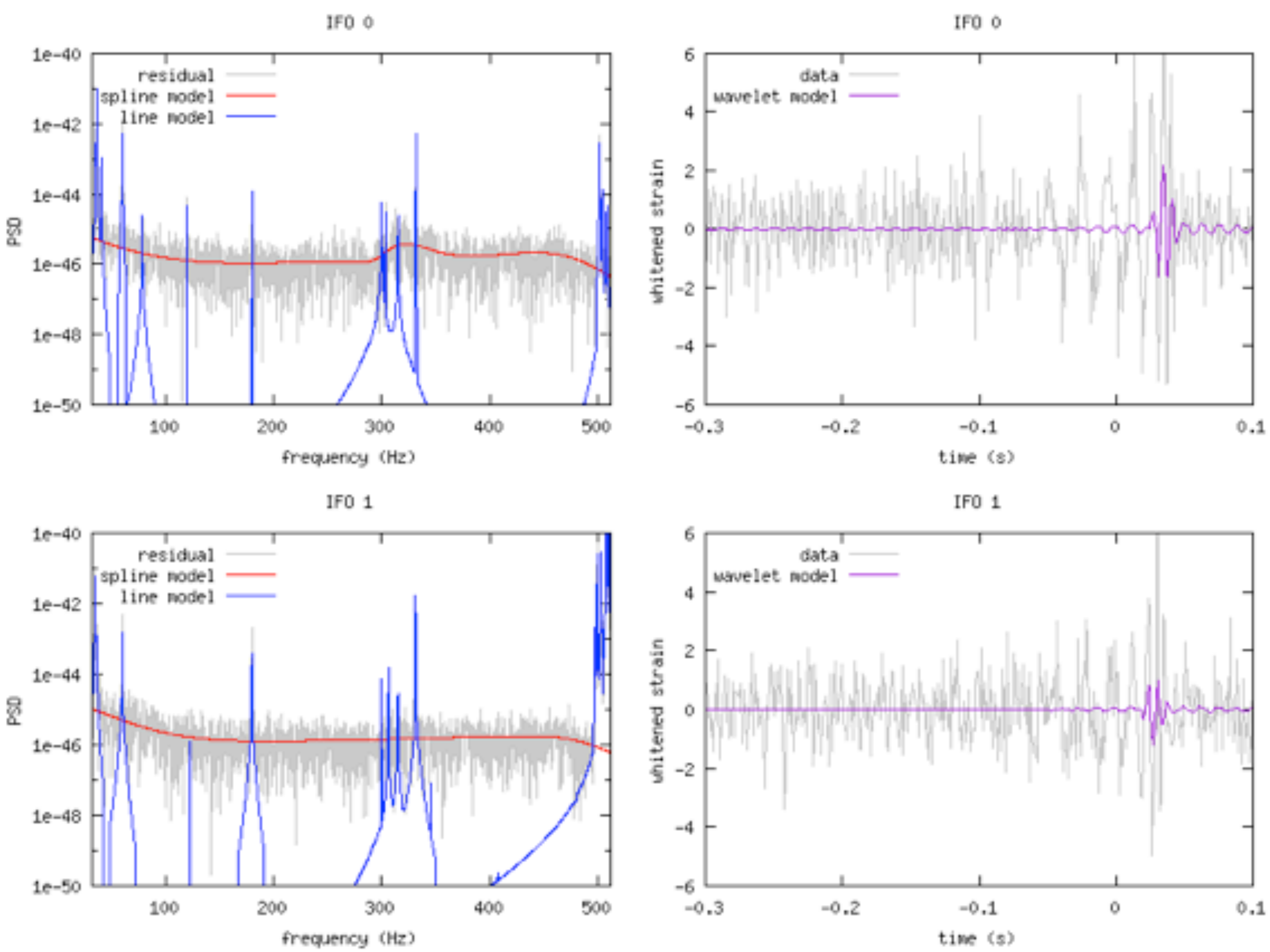
arxiv:1602.05411

Bayesian Noise Modeling

BayesWave/BayesLine:
Trans-dimensional Reverse Jump Markov-Chain Monte Carlo (RJMCMC) with simultaneous power spectrum, line fitting, and coherent/incoherent signal analysis







Source: N. Cornish, M. Millhouse (Montana State) T. Littenberg (NASA/MSFC)

arxiv:1410.3852
arxiv:1410.3835

Desiderata / Topics Skipped / In Progress

- Weakly-model dependent transient searches, etc...
- Hybrid Monte-Carlo / direct posterior gridding
- Model selection in tests of general relativity
- More advanced methods of hierarchical modeling
  - mass and spin distributions
  - model dependent formation channel determination
- Noise fitting / removal / classification
  - principle component analysis, machine/deep learning, Gaussian process regression