Testing General Relativity with Gravitational Waves from Coalescing Compact Binaries in the Advanced Detector Era

Michalis Agathos

Jeroen Meidam, Chris Van Den Broeck, Tjonnie Li,
Walter Del Pozzo, Salvatore Vitale, John Veitch

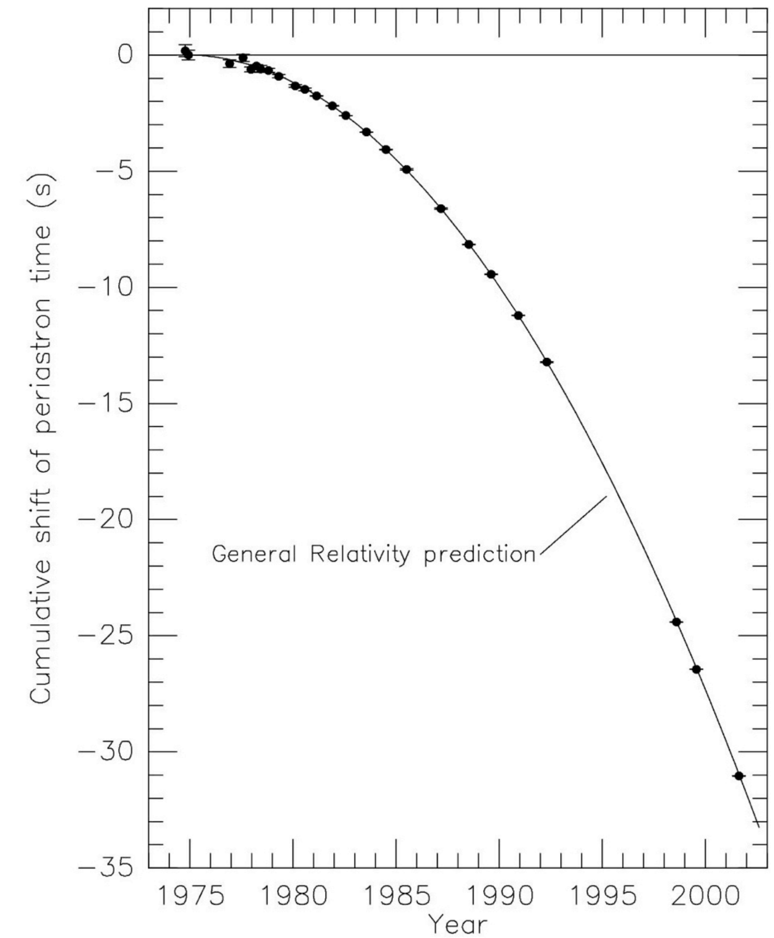
Phys. Rev. D 85 (2012), 082003 [arXiv:1110.0530]
J.Phys.Conf.Ser. 363 (2012) 012028 [arXiv:1111.5274]
MG13 Proceedings (2012) [arXiv:1305.2963]
Phys. Rev. D 89 (2014), 082001 [arXiv:1311.0420]

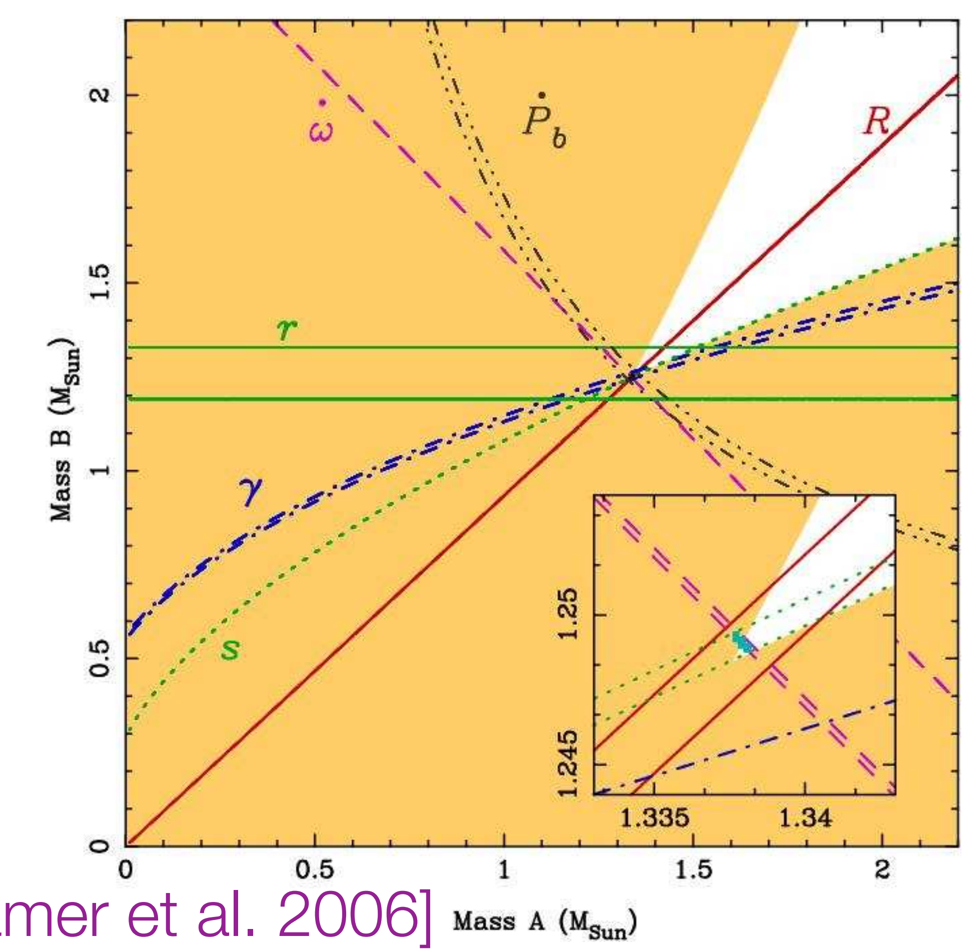
8th Aegean summer school
Rethymno, 30 June 2015

Testing GR: Motivation

- GR has passed all tests to date
- **But:** up to now, all tests were in weak field regime and/or involving slow-moving sources
- Binary Pulsar: $M/R \sim 10^{-6}$, $v/c \sim 10^{-3}$
- Gravitational waves from CBC will probe strong field gravity ($M/R \sim 0.2$) and relativistic sources ($v/c \sim 0.4$)
- These sources are perfect candidates for detection with aLigo/AdVirgo

[Weisberg & Taylor 2003]





[Kramer et al. 2006]

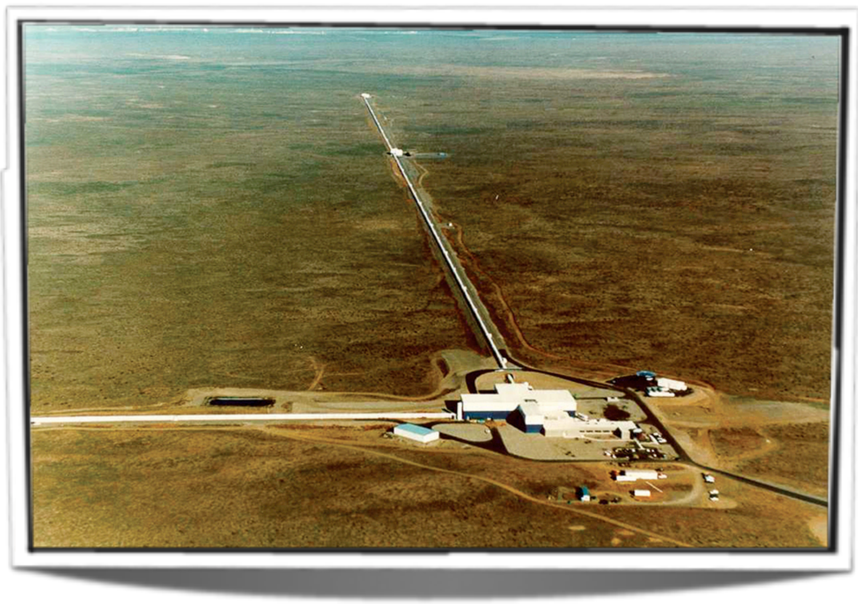
Advanced GW detectors

- The transition to the era of Advanced Interferometers (mid 2015) is expected to provide us with the first detections of GW signals

- Neutron Star binaries of M ~ a few Msun radiate within the *frequency bucket* of Adv LIGO/VIRGO

$$f_{lso} \sim 1600 \left(\frac{2.8\, M_\odot}{M_{\rm tot}} \right) \mathrm{Hz}$$

Advanced GW detectors



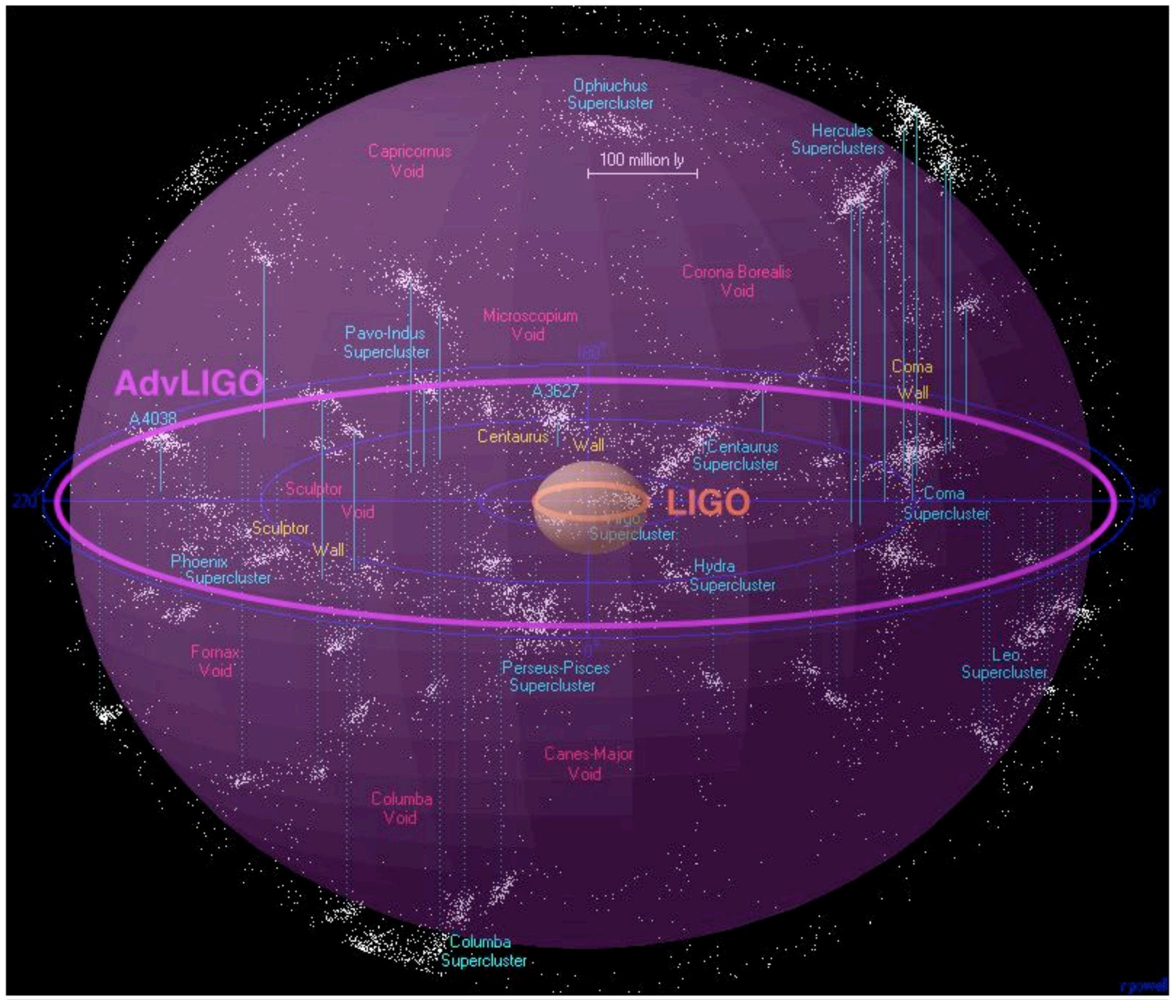
- The transition to the era of Advanced Interferometers (mid 2015) is expected to provide us with the first detections of GW signals

- Neutron Star binaries of M ~ a few Msun radiate within the *frequency bucket* of Adv LIGO/VIRGO

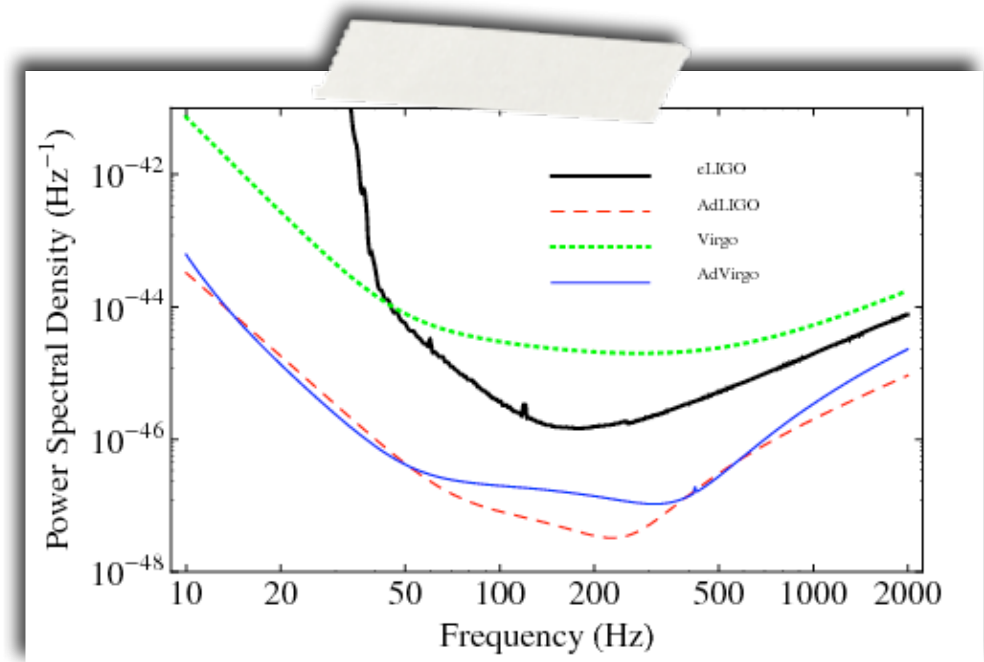
$$f_{lso} \sim 1600 \left(\frac{2.8\, M_{\odot}}{M_{\rm tot}} \right) \mathrm{Hz}$$

60 Mpc!

Expected rates for advanced detectors

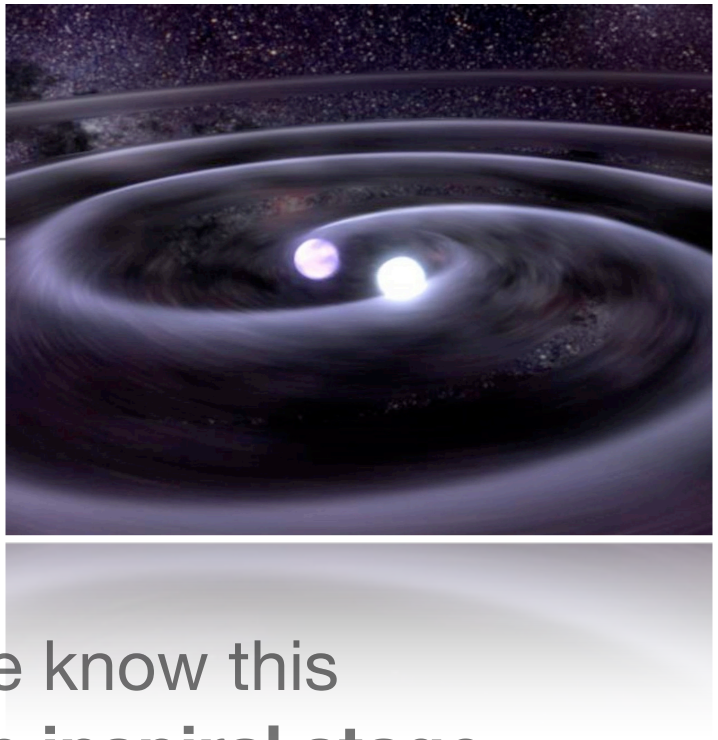


Network	Source	$\dot{N}_{\mathrm{low}}$ (yr^{-1})	$\dot{N}_{\mathrm{re}}$ (yr^{-1})	$\dot{N}_{\mathrm{high}}$ (yr^{-1})
	NS-NS	2×10^{-4}	0.02	0.2
Initial	NS-BH	7×10^{-5}	0.0004	0.1
	BH-BH	2×10^{-4}	0.007	0.5
	NS-NS	0.4	40	400
Advanced	NS-BH	0.2	10	300
	BH-BH	0.4	20	1000



Advanced LIGO/Virgo upgrade improves sensitivity by O(10)

The Post-Newtonian approximation



- The PN expansion gives approximate solutions to the 2-body problem in GR
- We consider binary neutron star (BNS) systems where we know this approximation to be very accurate for the best part of the **inspiral stage**.
- Simple frequency-domain waveform (TaylorF2) for systems with no spins:

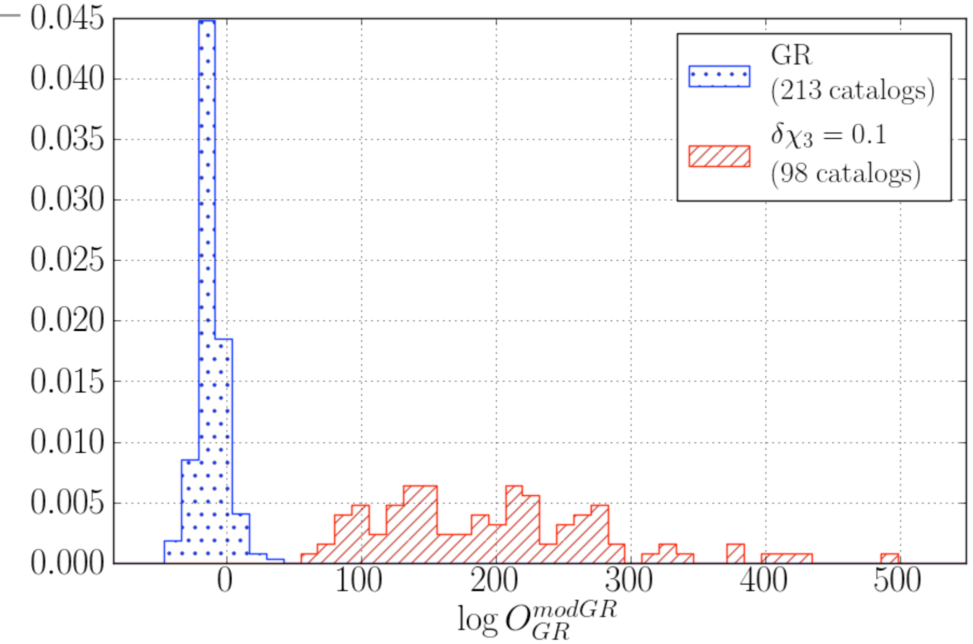
$$\tilde{h}(f) = \mathcal{A} f^{-7/6} \cos(2\Phi(f; m_1, m_2) + \phi_0) \quad v = (\pi M f)^{1/3}$$

$$\Phi = \left(\frac{v}{c}\right)^{-5} \sum_{i=0}^{N} \left[\psi_i + \psi_i^{(l)} \ln \frac{v}{c}\right] \left(\frac{v}{c}\right)^i$$

- Deviation from GR should give a different functional dependence of the PN phase coefficients $\psi_i (m_1, m_2, \mathbf{S}_1, \mathbf{S}_2)$, $\psi_i^{(l)} (m_1, m_2, \mathbf{S}_1, \mathbf{S}_2)$ on masses & spins

Testing General Relativity with GW

- If GR is violated the orbital evolution will be different than what GR predicts.



- With a few realistically quiet detections we will be able to test GR like never before!

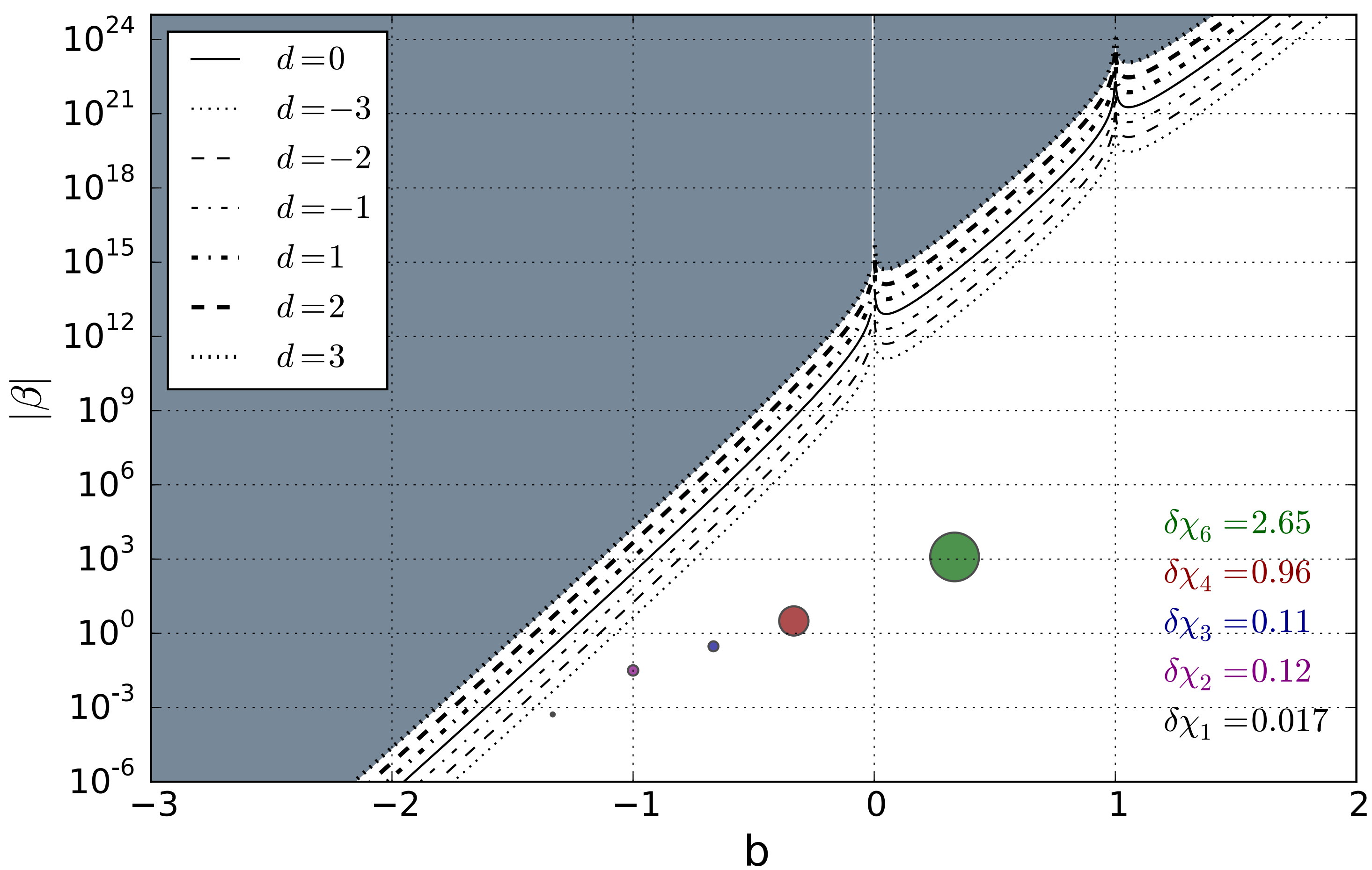
- e.g. consider a heuristic phase modification

$$\beta\eta^d(\pi\mathcal{M}_c f)^b$$

[Yunes & Hughes 2010]

- With single GW detection we can beat binary pulsar constraints by many orders of magnitude

  Li et al 2011 [arXiv:1110.0530]
  Li et al 2011 [arXiv:1111.5274]
  Agathos et al 2013 [arXiv:1311.0420]
  Agathos et al 2013 [arXiv:1305.2963]

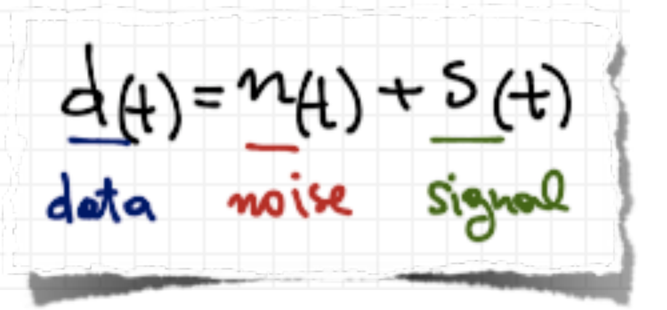


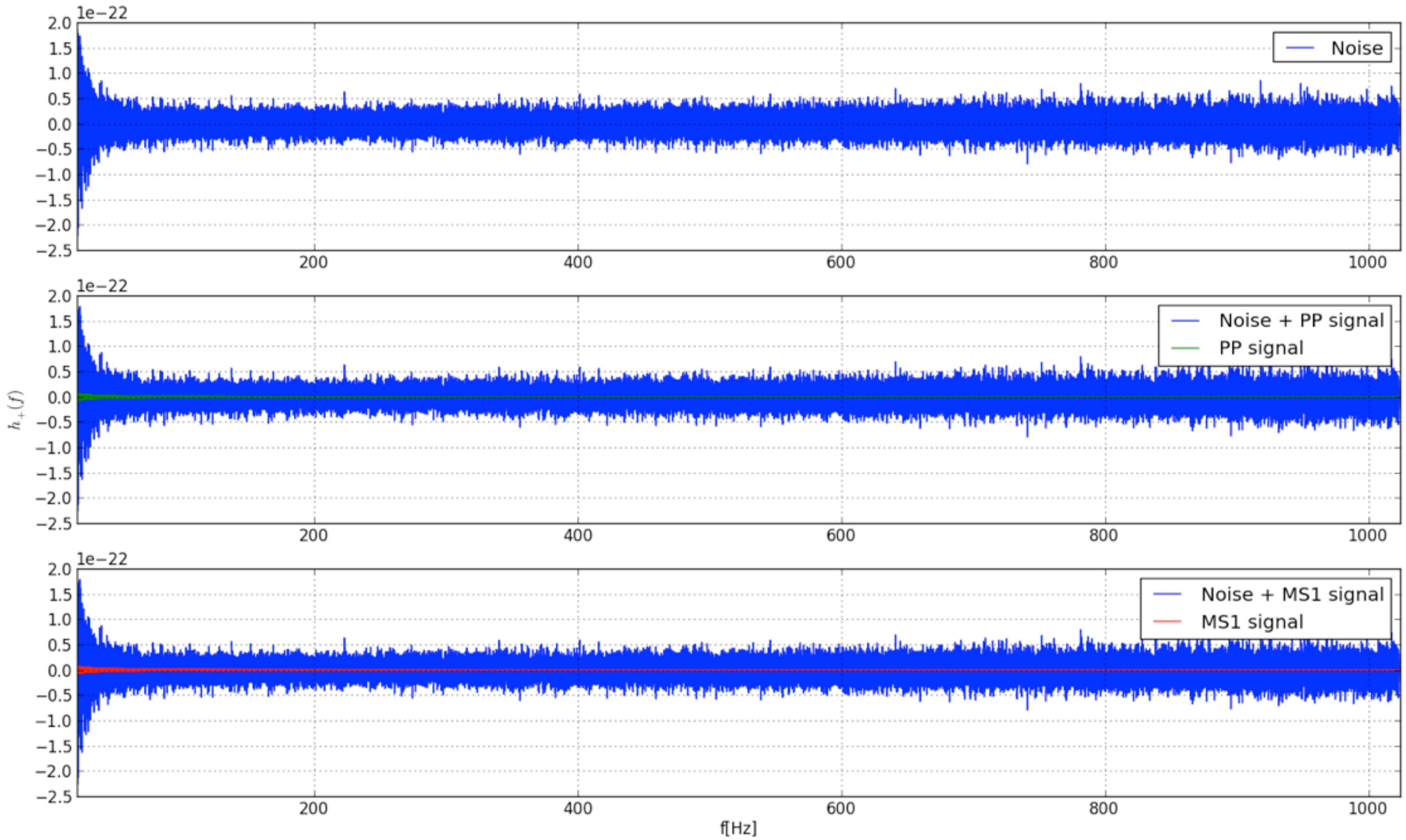
Desiderata

- Need a “theory independent” test for GR, not based on a particular alternative. If GR is violated, it may be in a way that has not yet been envisaged.
- Should be as generic as possible, parametrizable and computationally feasible.
- Has to be reliable for “quiet” sources.
- Should have ability to combine information from multiple sources, so as to arrive at a more stringent test of GR.
- Should not be tied to a particular waveform approximant.

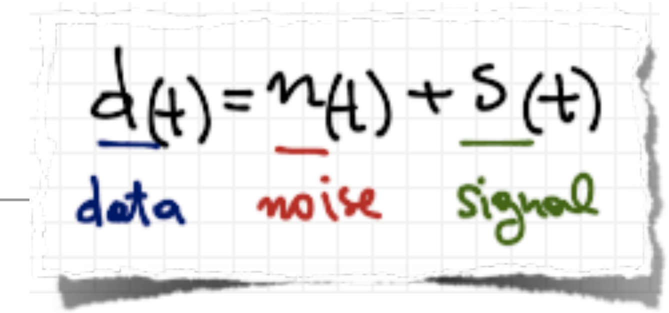
GW data analysis

- Data is dominated by noise
- Need to use stochastic properties of noise to dig out signal





GW data analysis



- Signal recovery from noisy data is possible because:
  - we know how the stochastic properties of our noise

$$\langle \tilde{n}^*(f)\tilde{n}(f')\rangle = \frac{1}{2}\delta(f-f')S_n(f)$$

  - we know how a signal should look like
  - each signal gives many data points

- Signal recovery is terribly complicated because:
  - there are many (~15) free parameters that define each source
  - noise can often mimic a GW (false alarm)
  - GR equations are hard to solve accurately

- Parameters for compact binary sources: $\{m_1, m_2, \mathbf{S}_1, \mathbf{S}_2, D, \theta, \phi, \iota, \psi, t_c, \phi_c\}$

- Noise-weighted inner product for matched filtering: $\langle a, b\rangle \equiv \Re\left\{\int_{-\infty}^{\infty} \mathrm{d}f\, \frac{\tilde{a}^*(f)\tilde{b}(f)}{\frac{1}{2}S_n(f)}\right\}$

$$P(A|B)P(B) = P(B|A)P(A)$$

Bayesian inference in GW data analysis

- CBC waveforms encompass a high-dimensional parameter space

- We want to calculate the *evidence* of the model hypothesis by marginalizing the *likelihood* over the parameter space:

$$P(d|H,I) = \int \mathrm{d}\vec{\theta}\, p(\vec{\theta}|H,I)\, p(d|\vec{\theta},H,I)$$

- Also interested in posterior PDF for parameter estimation: $p(\vec{\theta}|d,H,I)$

- Efficiently sample parameter space and obtain both evidence and posterior using the *Nested Sampling* algorithm [Skilling 2006] [Veitch & Vecchio 2009]

- Combine information from independent events:

$$p(d_1,d_2|H,I) = p(d_1|H,I)p(d_2|H,I)$$

Bayesian inference in GW data analysis

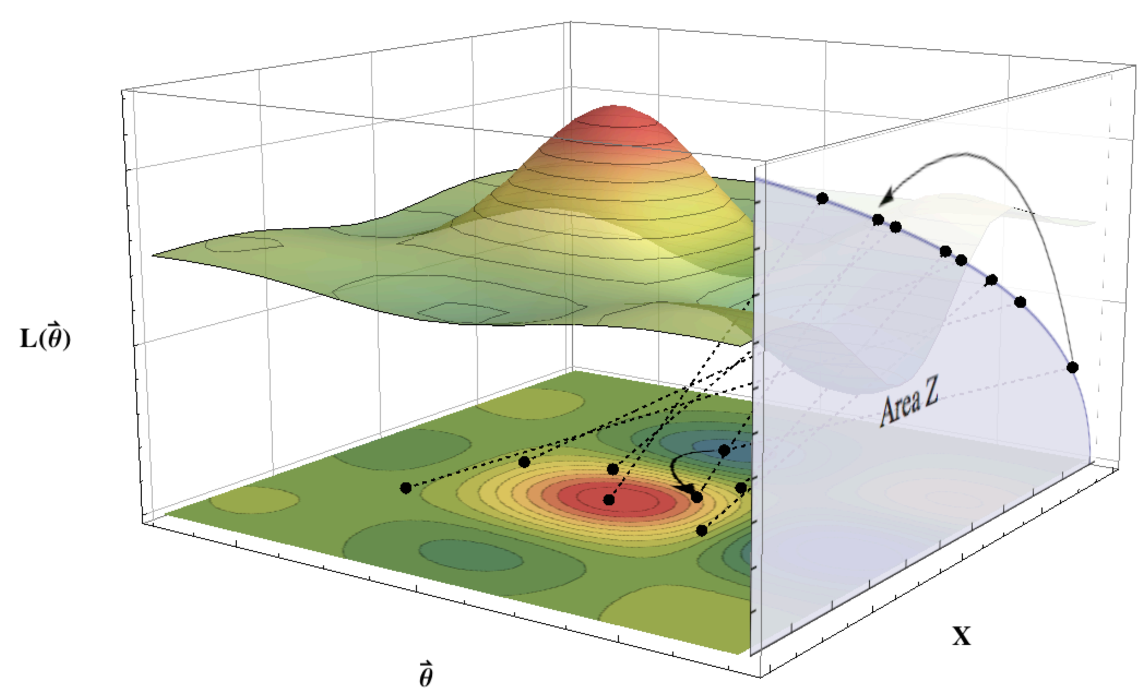
- Likelihood:

$$p(d|\vec{\theta}, H, I) = p(d - h_{\vec{\theta}}|H_{noise}, I) = \mathcal{N}\ \exp\left[-\int_{-\infty}^{\infty} \mathrm{d}f\ \frac{|\tilde{d}(f) - \tilde{h}(f;\vec{\theta})|^2}{S_n(f)}\right]$$

$$= \mathcal{N}\ e^{-\frac{\langle d - h_{\vec{\theta}}, d - h_{\vec{\theta}}\rangle}{2}}$$

- Nested sampling explores the parameter space by climbing up the likelihood function

- Evidence is numerically accumulated



Bayesian Model Selection

$$P(A|B)\,P(B) = P(B|A)\,P(A)$$

- For competing hypotheses, given a set of data d, calculate the posterior probability for each hypothesis:

$$P(H_i|d,\mathrm{I}) = \frac{\overbrace{P(d|H_i,\mathrm{I})}^{\text{evidence}}\,\overbrace{P(H_i|\mathrm{I})}^{\text{prior}}}{P(d|\mathrm{I})}$$

- Define the **odds ratio** between a pair of hypotheses:

$$\mathcal{O}_2^1 = \frac{P(H_1|d,\mathrm{I})}{P(H_2|d,\mathrm{I})} = \frac{P(H_1|\mathrm{I})}{P(H_2|\mathrm{I})}\,\frac{P(d|H_1,\mathrm{I})}{P(d|H_2,\mathrm{I})}$$

- Combine information from multiple sources:

$$\mathcal{O}_2^1 = \frac{P(H_1|d_1,\ldots,d_n,\mathrm{I})}{P(H_2|d_1,\ldots,d_n,\mathrm{I})} = \frac{P(H_1|\mathrm{I})}{P(H_2|\mathrm{I})}\prod_{i=1}^{n}\frac{P(d_i|H_1,\mathrm{I})}{P(d_i|H_2,\mathrm{I})}$$

TIGER
(**T**est **I**nfrastructure for **GE**neral **R**elativity)

- Define the GR and modGR hypotheses for the phase evolution:

> **H$_{GR}$** : All PN phase coefficients have the functional dependence on the masses (and spins) that is predicted by GR
>
> **H$_{modGR}$** : One or more of the phase coefficients are **not** as predicted by GR, without specifying which $\psi_i = \psi_i^{GR}(1 + \delta\chi_i)$

- Reformulate H$_{modGR}$ as union of testable disjoint sub-hypotheses:

$$H_{modGR} = \bigvee \mathcal{H}_{i_1 \dots i_k} \qquad P(H_{modGR}|d, \mathrm{I}) = \sum_{k; i_1 < \dots < i_k} P(\mathcal{H}_{i_1 \dots i_k}|d, \mathrm{I})$$

$\mathcal{H}_{i_1 \dots i_k}$ is the model where $\{\psi_{i_1}, \cdots, \psi_{i_k}\}$ are free to deviate away from GR, but not the others

- Combine information from multiple sources to catalog's odds ratio

$$\mathcal{O}_{GR}^{modGR} = \frac{\alpha}{2^{N_T} - 1} \sum_{i_1 < \ldots < i_k; k \leq N_T} \prod_{A=1}^{\mathcal{N}} \frac{P(\mathrm{d}_A|H_{i_1 \ldots i_k}, \mathrm{I})}{P(\mathrm{d}|\mathcal{H}_{GR}, \mathrm{I})}$$

Simulations

1. **Simulated signals**:

    Generate simulated GR/modGR signals from **BNS** inspiral and simulate detector response and noise for 3 advanced detectors (HLV)

2. **Recovery**:

    Estimate **evidence** for the GR and modGR hypotheses for each source, by integrating the likelihood over the parameter space using the *nested sampling* algorithm

3. **Post-process**:

    Combine evidence from multiple sources to cumulative odds ratio between GR and modGR for **catalogues of sources**

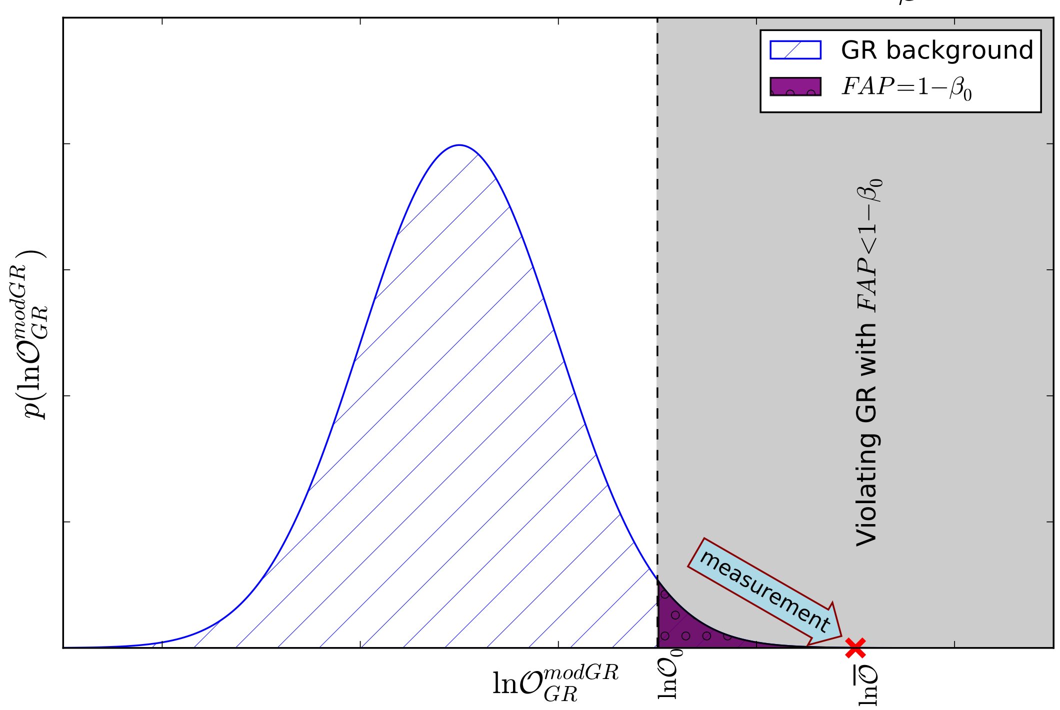
    The set of GR injections is used to form the *background* distribution of the statistic, within which deviations from GR are not measurable

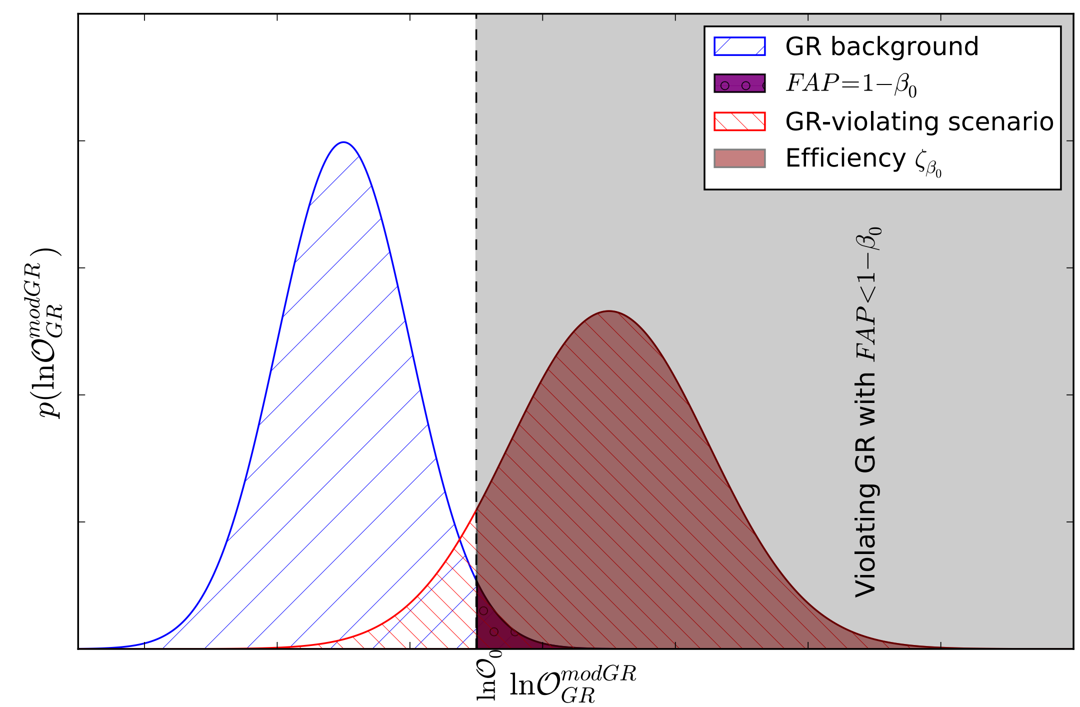
Background, foreground and efficiency

- The threshold for concluding that GR is violated is **not** at $O_{GR}^{modGR} = 1$!

- ***Efficiency***: how much of the foreground is above a given fraction of the background?

$$\zeta = \int_{\ln \mathcal{O}_\beta}^{\infty} P(ln\mathcal{O}|\kappa, \mathcal{H}_{non-GR}, I)\mathrm{d}\ln\mathcal{O}$$



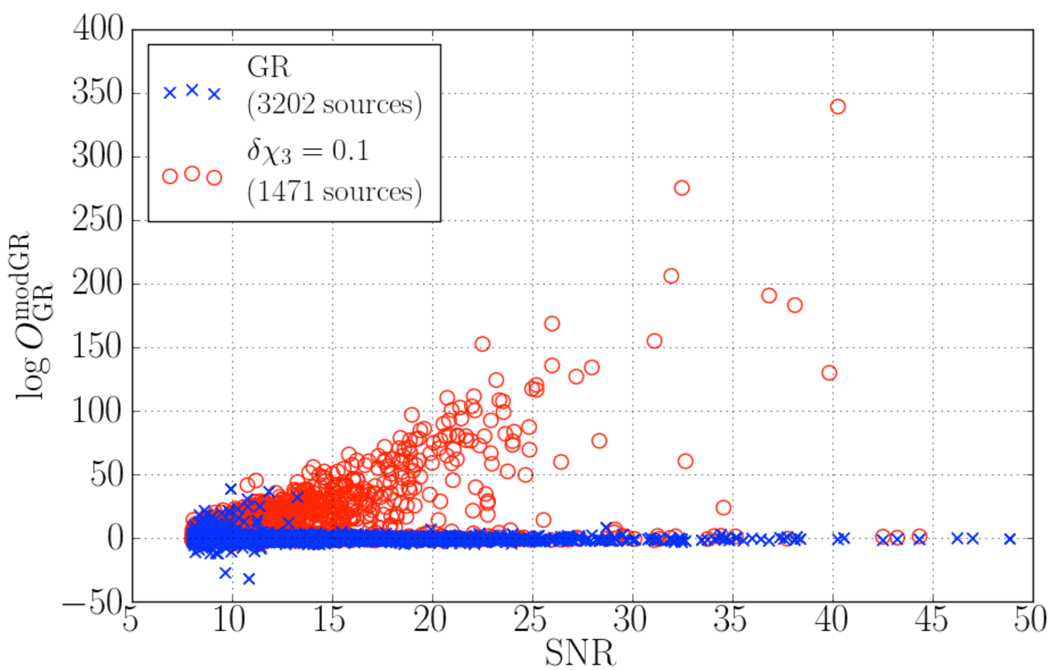


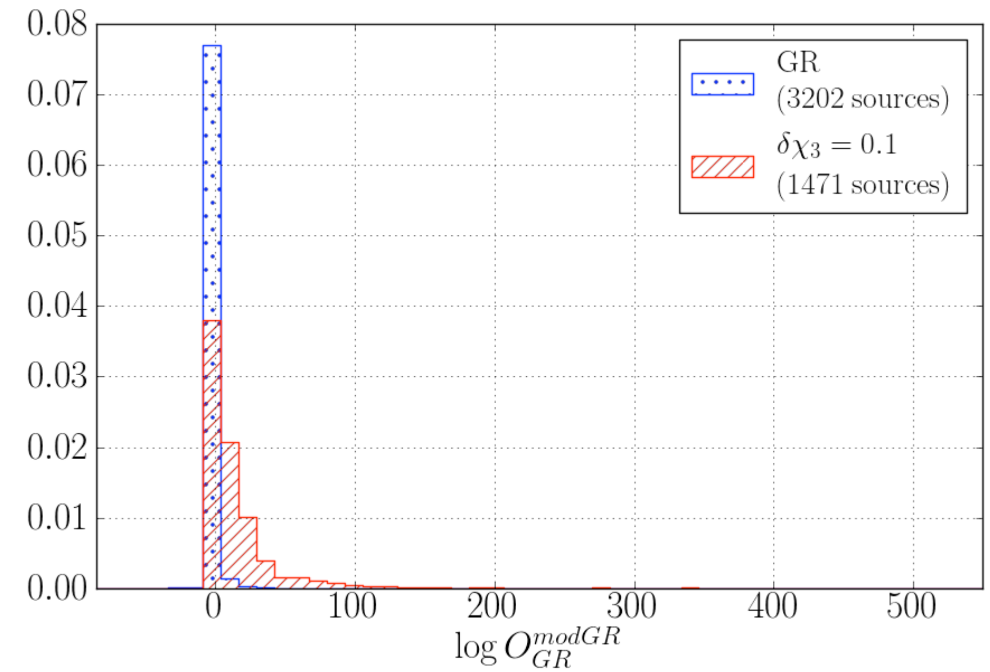
- where **β** is a given FAP

$$\beta = \int_{\ln \mathcal{O}_\beta}^{\infty} P(ln\mathcal{O}|\kappa, \mathcal{H}_{GR}, I)\mathrm{d}\ln\mathcal{O}$$

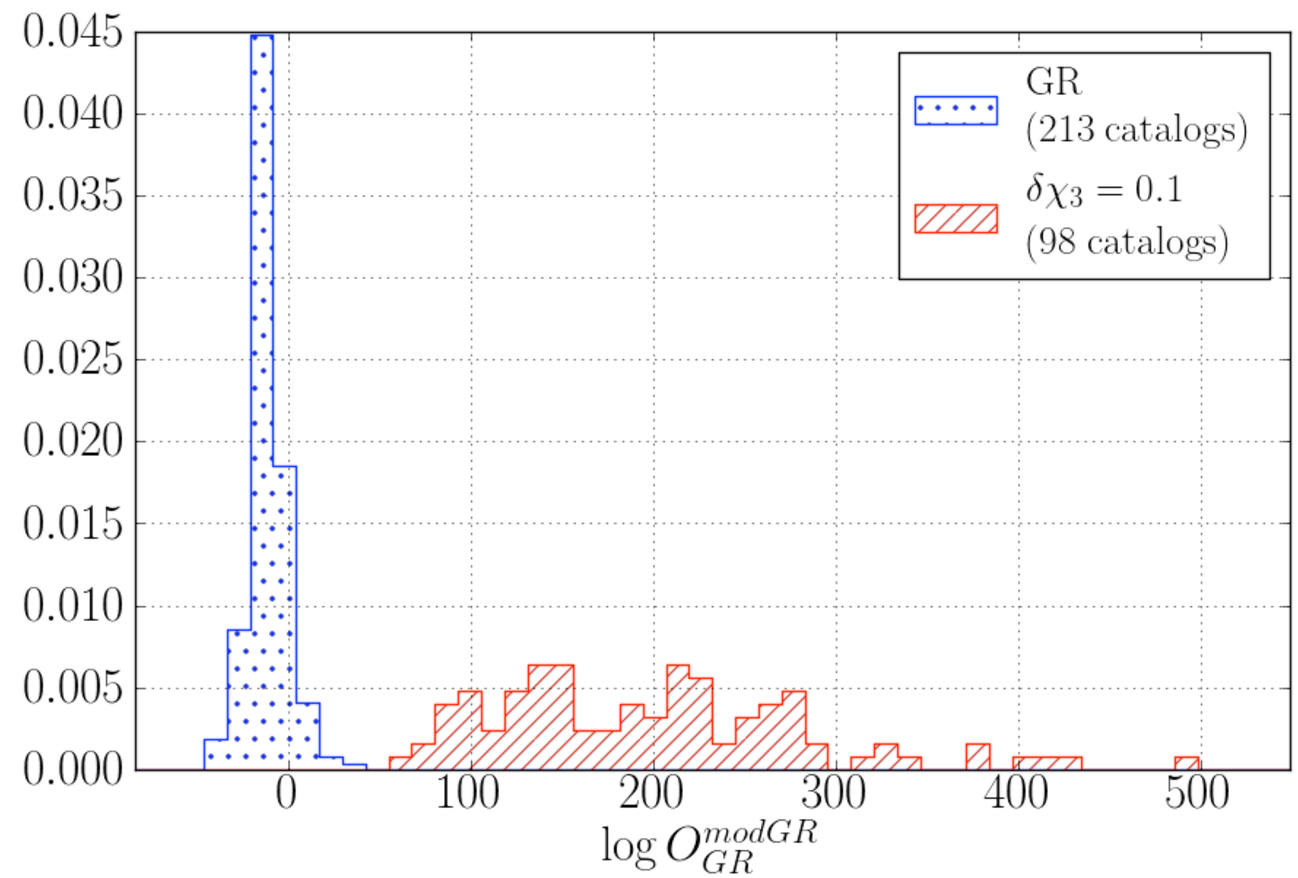
Li et al 2011 [arXiv:1110.0530]

First results : 10% shift at 1.5PN



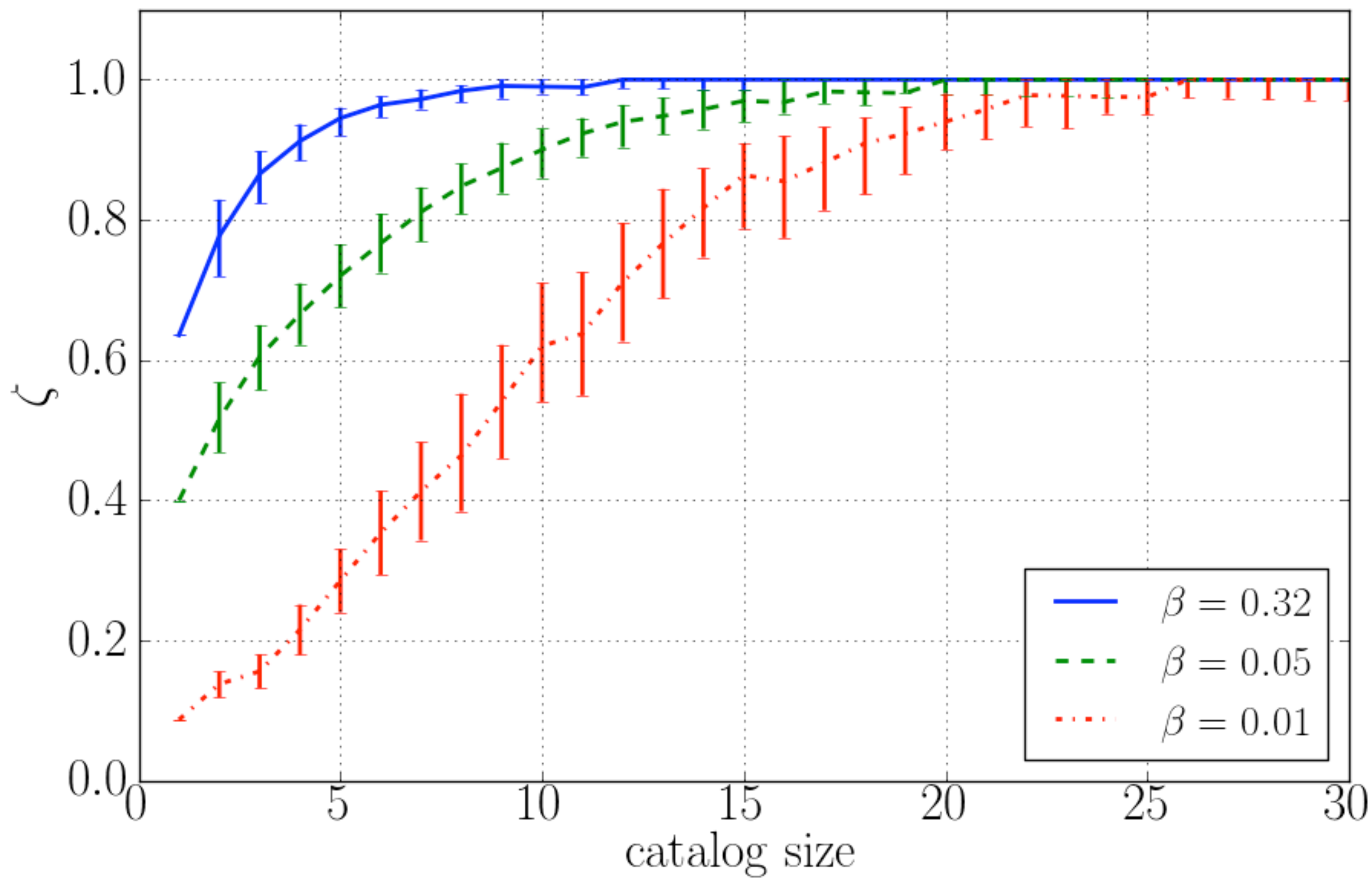


For a constant shift of 10% at 1.5PN, the efficiency is close to 100% for catalogs of 15 sources each



Efficiency for 10% shift at 1.5PN

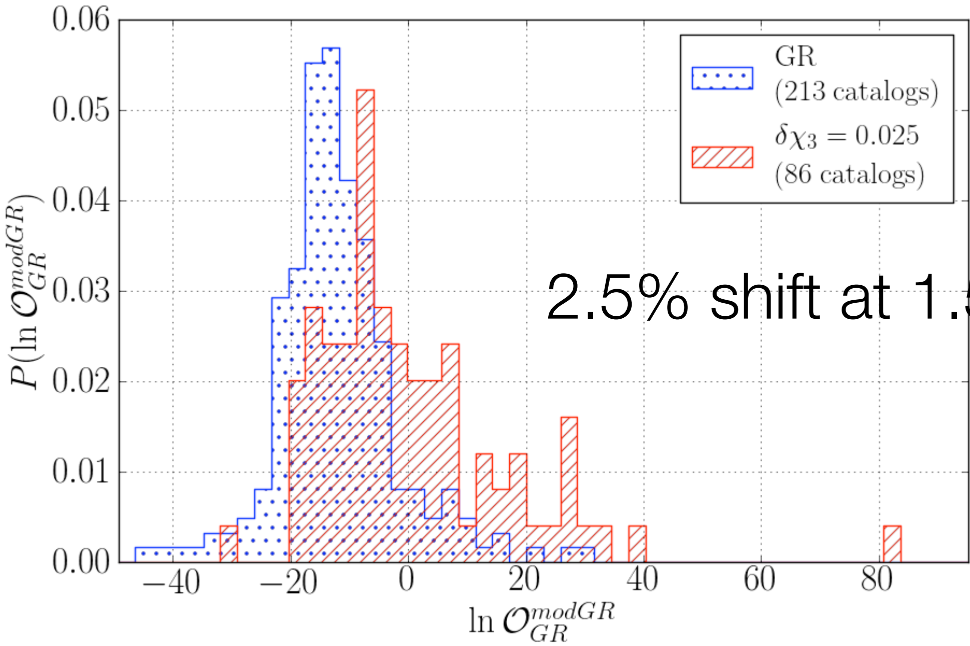
- How does our performance go with an increasing number of sources per catalog?



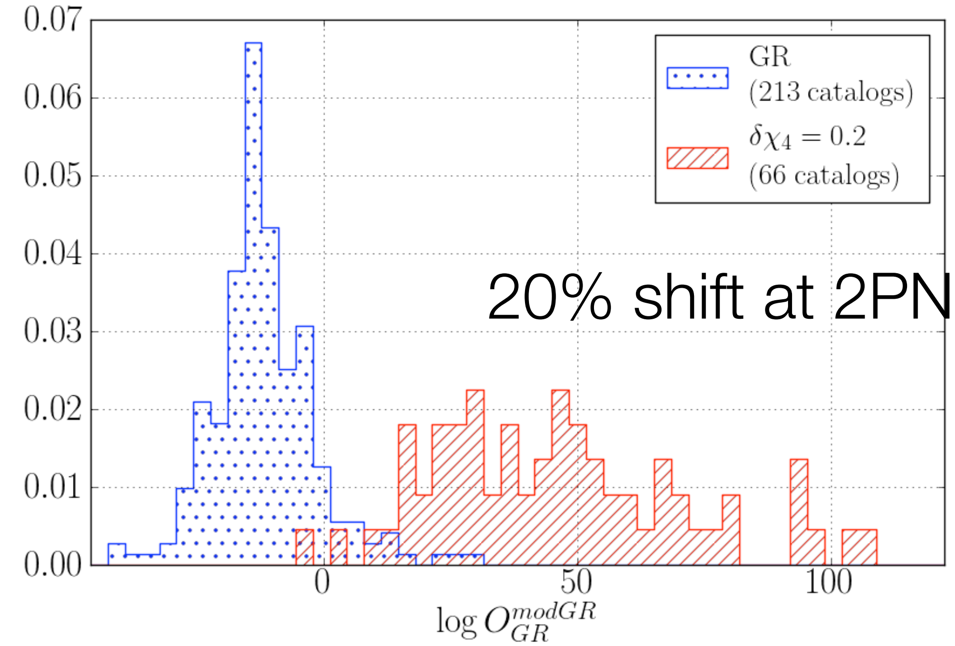
Testing different types of GR-violations

Li et al 2011 [arXiv:1110.0530]

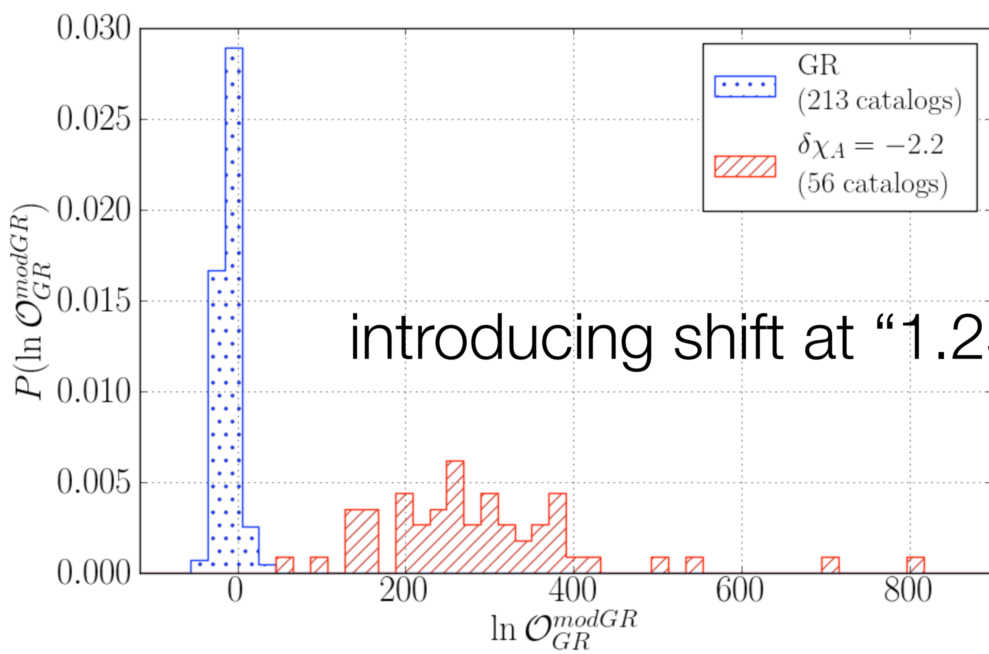
Li et al 2011 [arXiv:1111.5274]



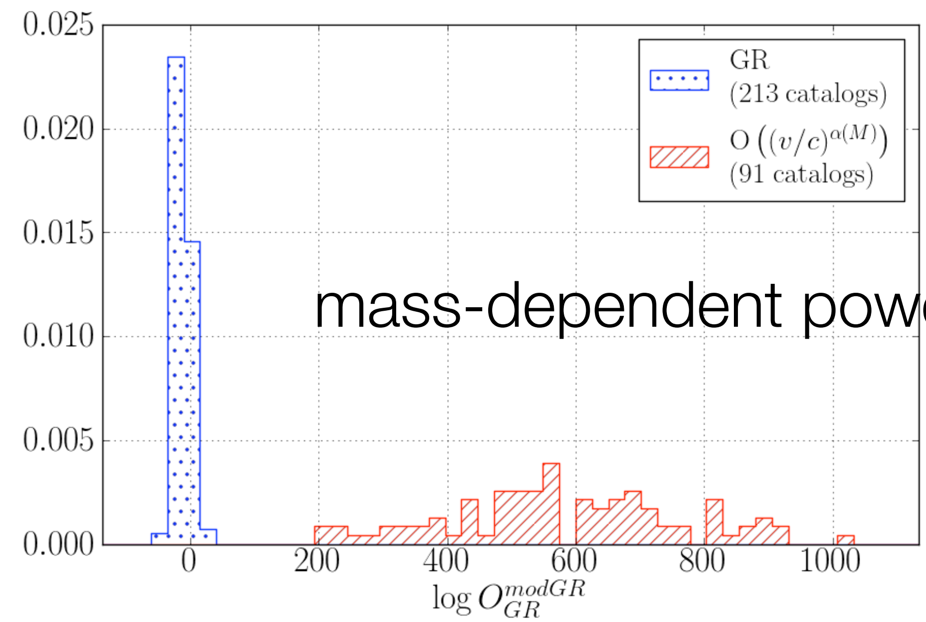
2.5% shift at 1.5PN



20% shift at 2PN



introducing shift at “1.25PN”



mass-dependent power of v/c

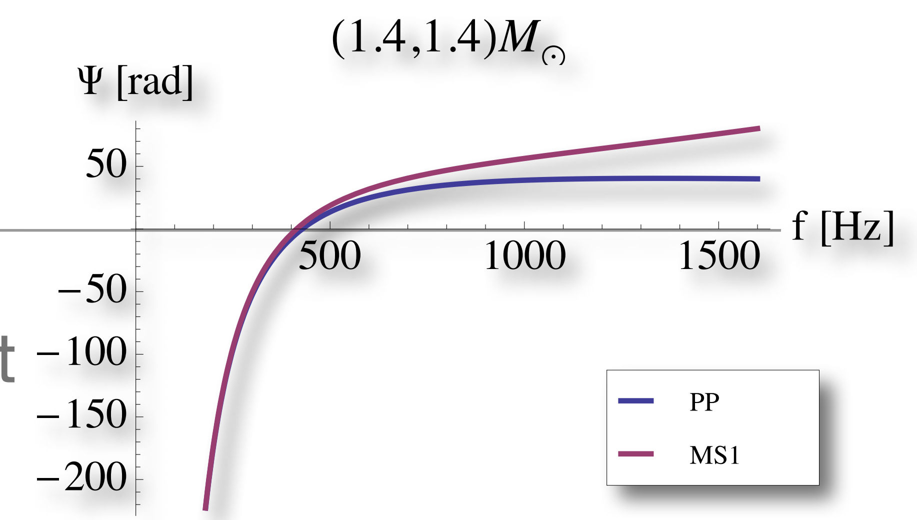
$$\Psi^{\rm GR}(\mathcal{M},\eta;f) \rightarrow \Psi^{\rm GR}(\mathcal{M},\eta;f) + \frac{3}{128\eta}(\pi M f)^{-2+M/(3M_\odot)}$$

Robustness

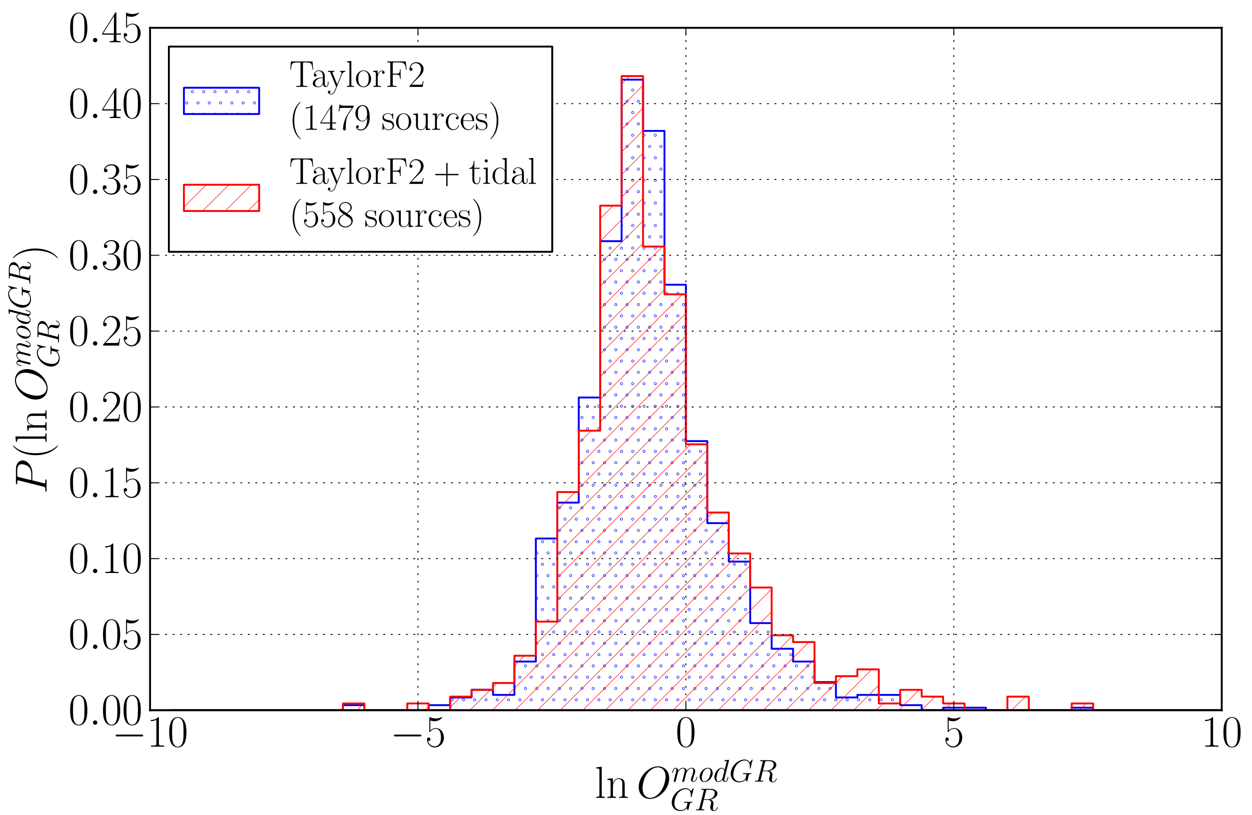
Agathos et al 2013 [arXiv:1311.0420]

- Tidal effects of unknown magnitude
- Effect of instrumental calibration errors
- Waveform mismatch, truncated PN expansion
- Effect of spins (aligned or generic)
- Real (non-gaussian, non-stationary) noise

Tidal effects

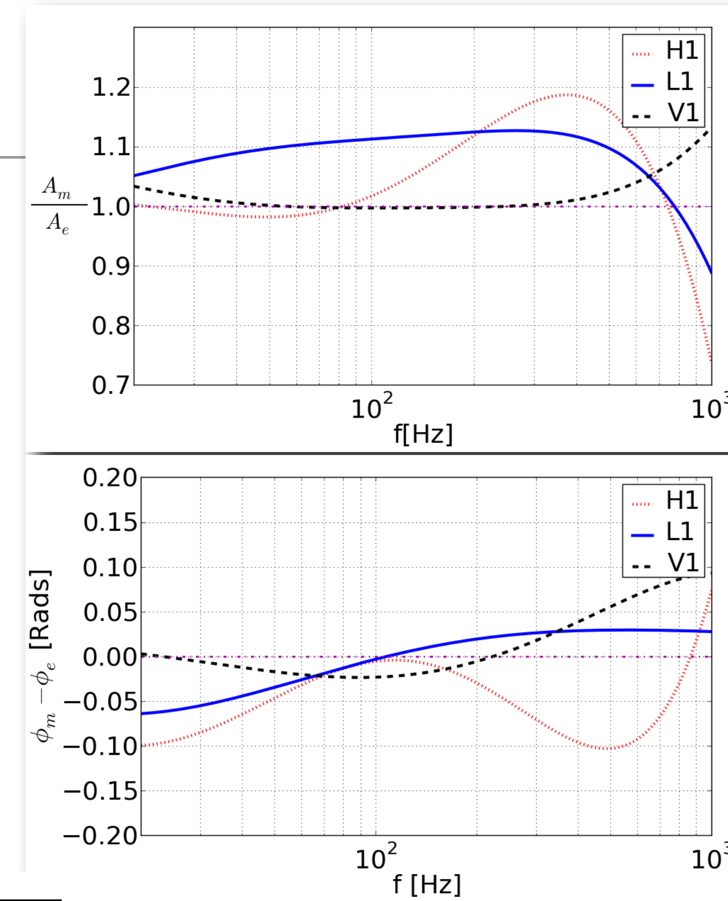


- For BNS systems, tidal effects will become important at high frequencies (>450Hz) $\Psi = \Psi_{PP} + \Psi_{tidal}$
  [Hinderer et al 2010]
- We **cut off** our analysis at 400Hz with a minor SNR loss ~1%
- Turn on tidal effects and see how performance is affected

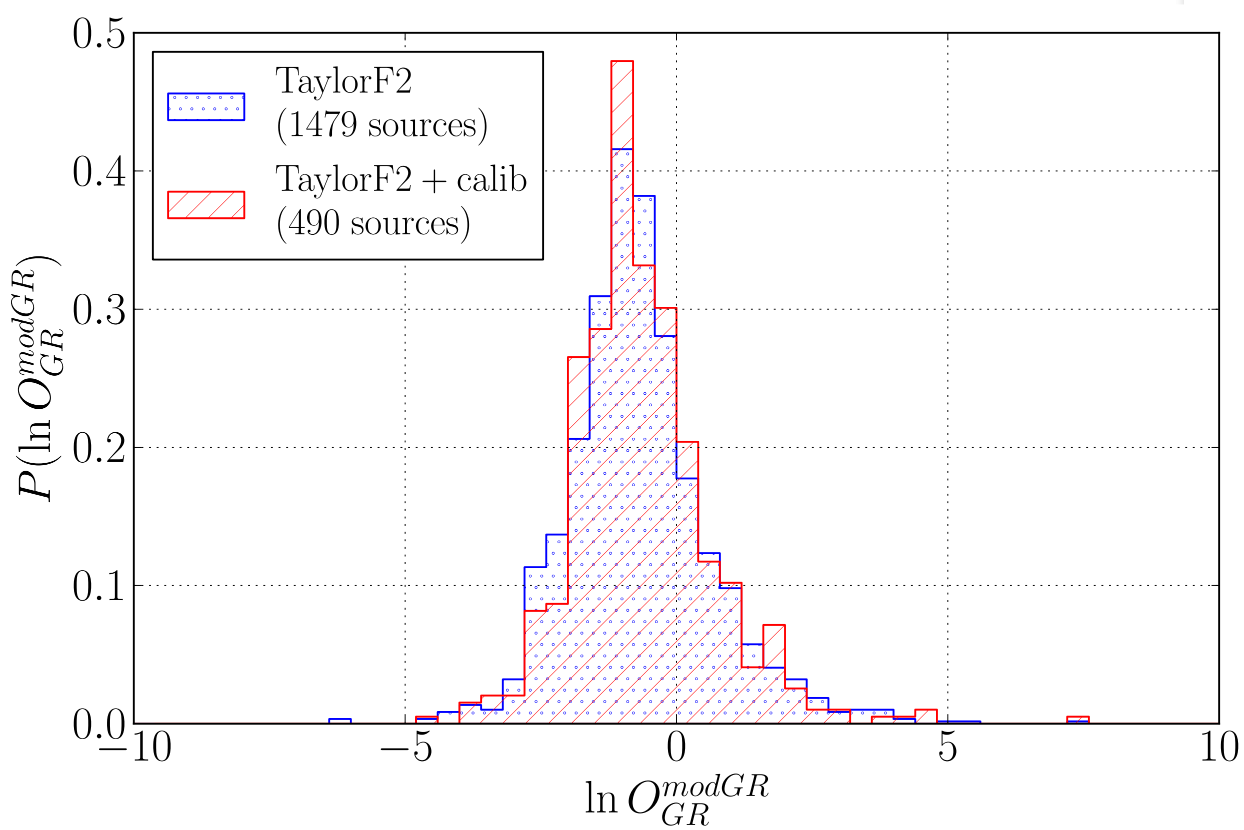


Vitale et al 2011 [arXiv:1111.3044]

Instrumental calibration errors



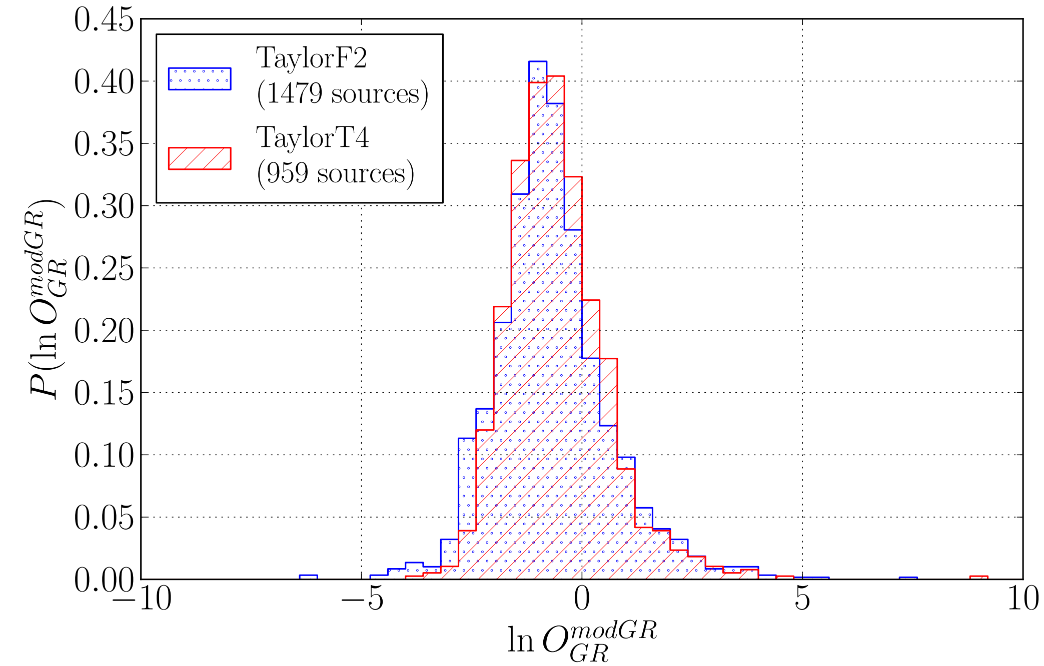
- Errors in the calibration of one or more parts of the detector may lead to “misinterpretation” of the data.
- No severe impact on parameter estimation
- What is the impact on TIGER?



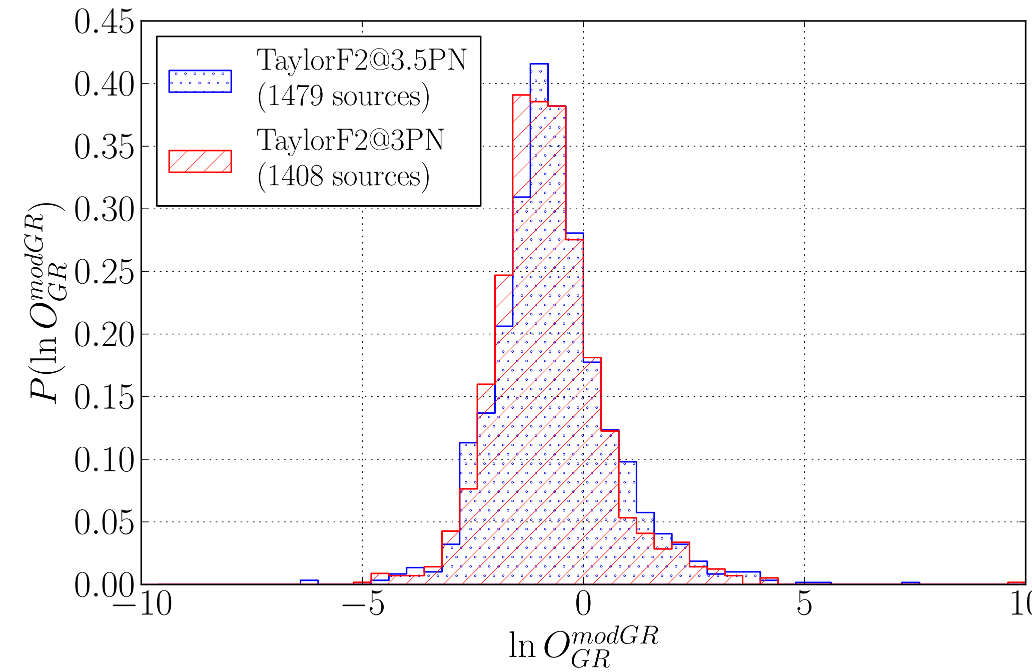
Waveform mismatch/missing PN terms

- For the BNS part of the parameter space (low masses) our waveform models are reliable and their mismatch is tiny [Buonanno et al. 2009]

1) Pick two different waveform approximants for the **signals**, but in both cases stick to TaylorF2 for recovery



2) Use the same waveform approximant for the **signals** (TaylorF2) but with the highest available PN term missing in the recovery waveform



Fully precessing spins

- Spin contributions enter the phase at 1.5PN order and beyond

$$\beta = \frac{1}{12}\sum_{i=1}^{2}\left[113\left(\frac{m_i}{M}\right)^2 + 75\eta\right]\hat{L}\cdot\vec{S}_i$$

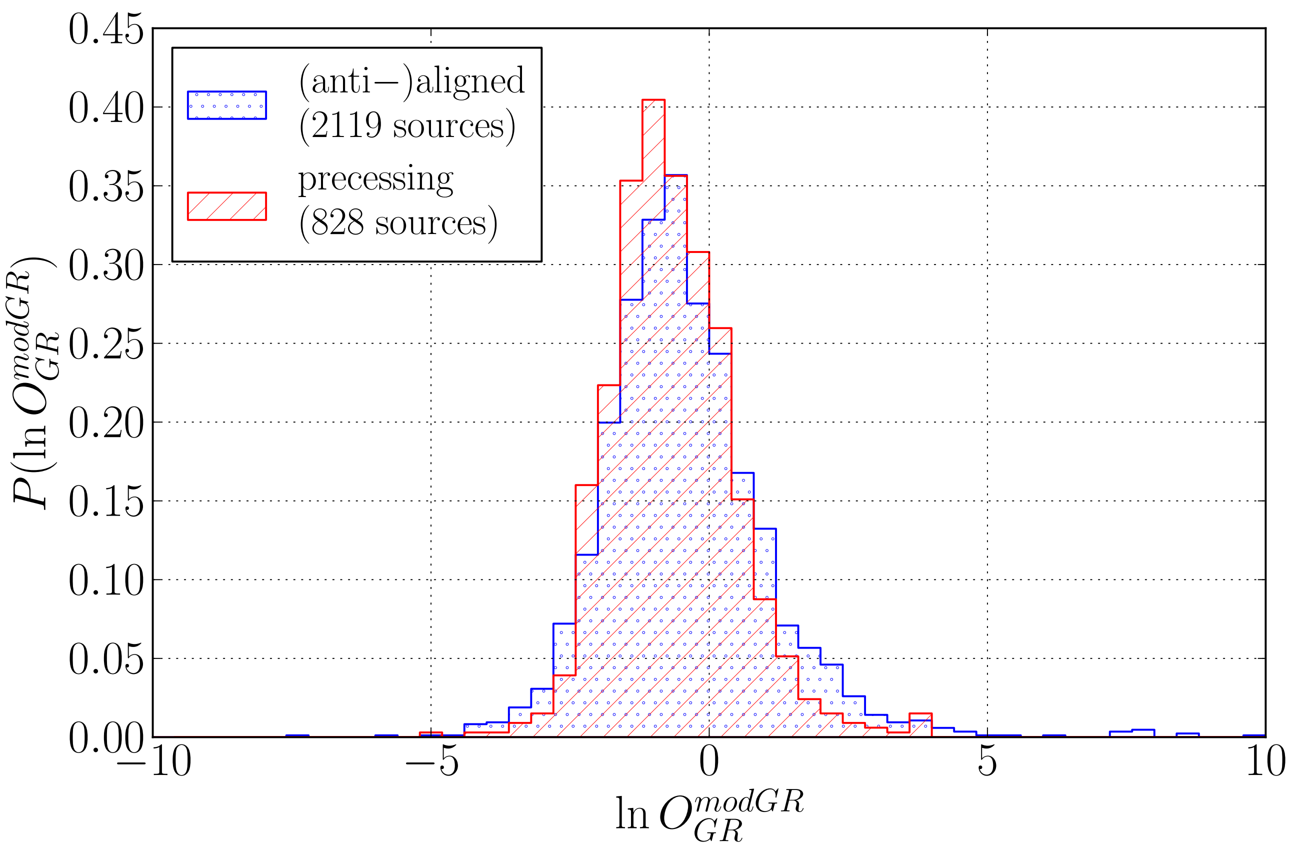
- From the known NS binaries [O'Shaughnessy & Kim 2009] we expect spins to be small

$$\psi_3 = [-16\pi + 4\beta](\pi M)^{-2/3}$$

[Kidder et al. 1993]

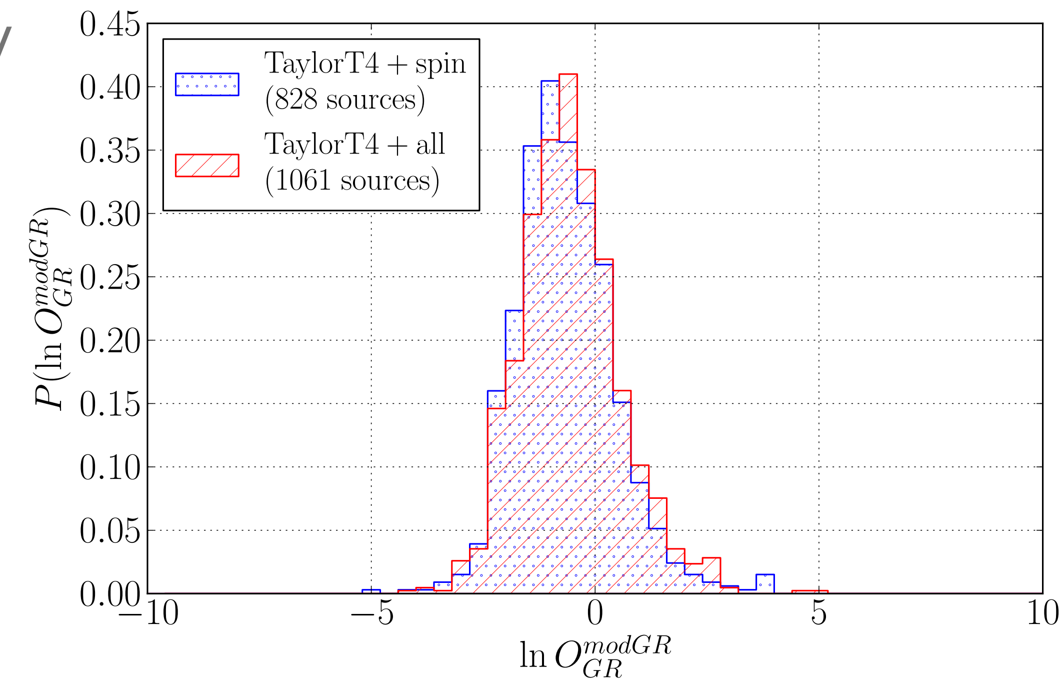
- Use **SpinTaylorT4** to simulate sources with spins of magnitude following a gaussian distribution (μ=0, σ=0.05) and of **random orientation**

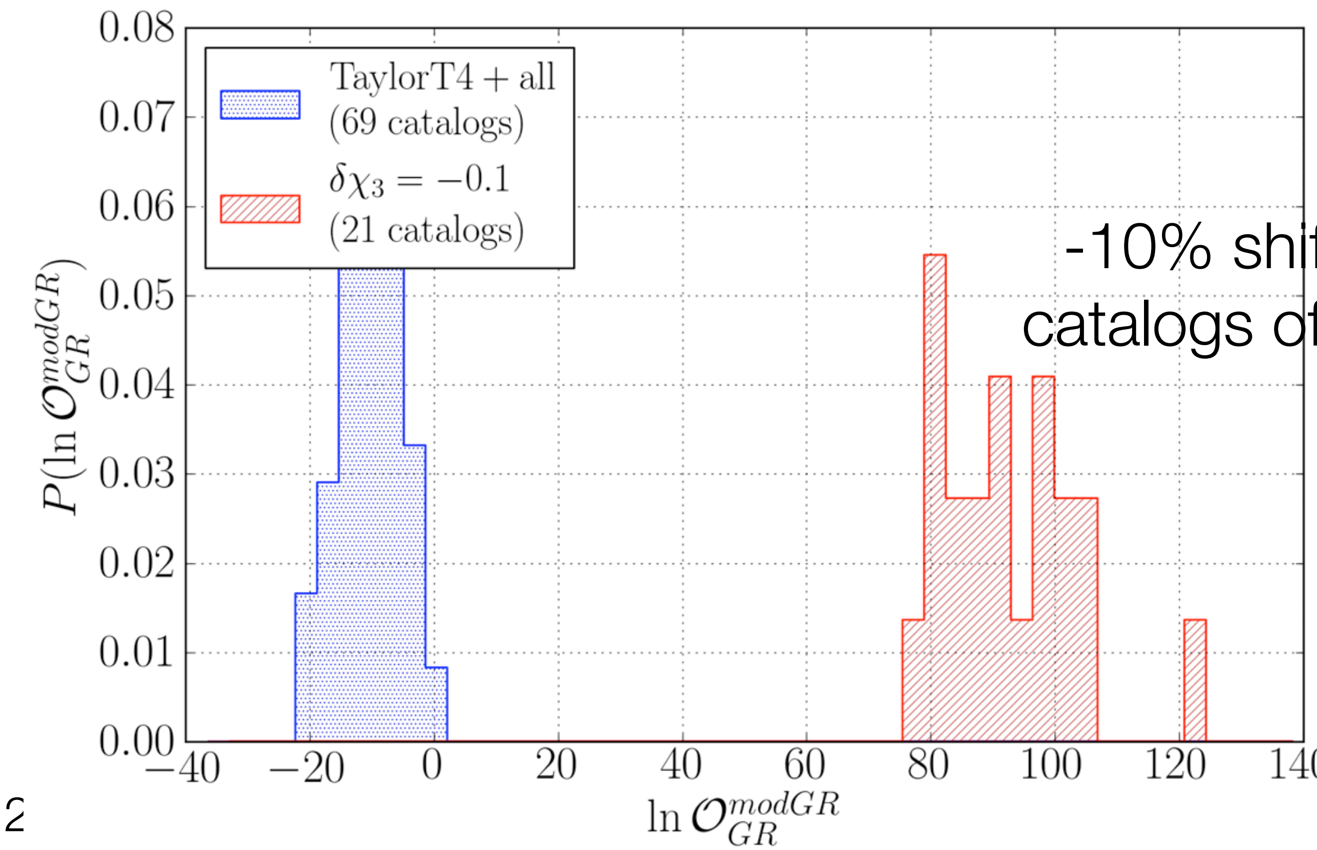
- Recover with (anti-)aligned spinning TaylorF2



Putting everything together

- ✓ SpinTaylorT4 (different waveforms for injection/recovery)
- ✓ Tidal effects, 5PN + 6PN in phase using a very hard equation of state
- ✓ Generic, precessing spins
- ✓ Calibration errors

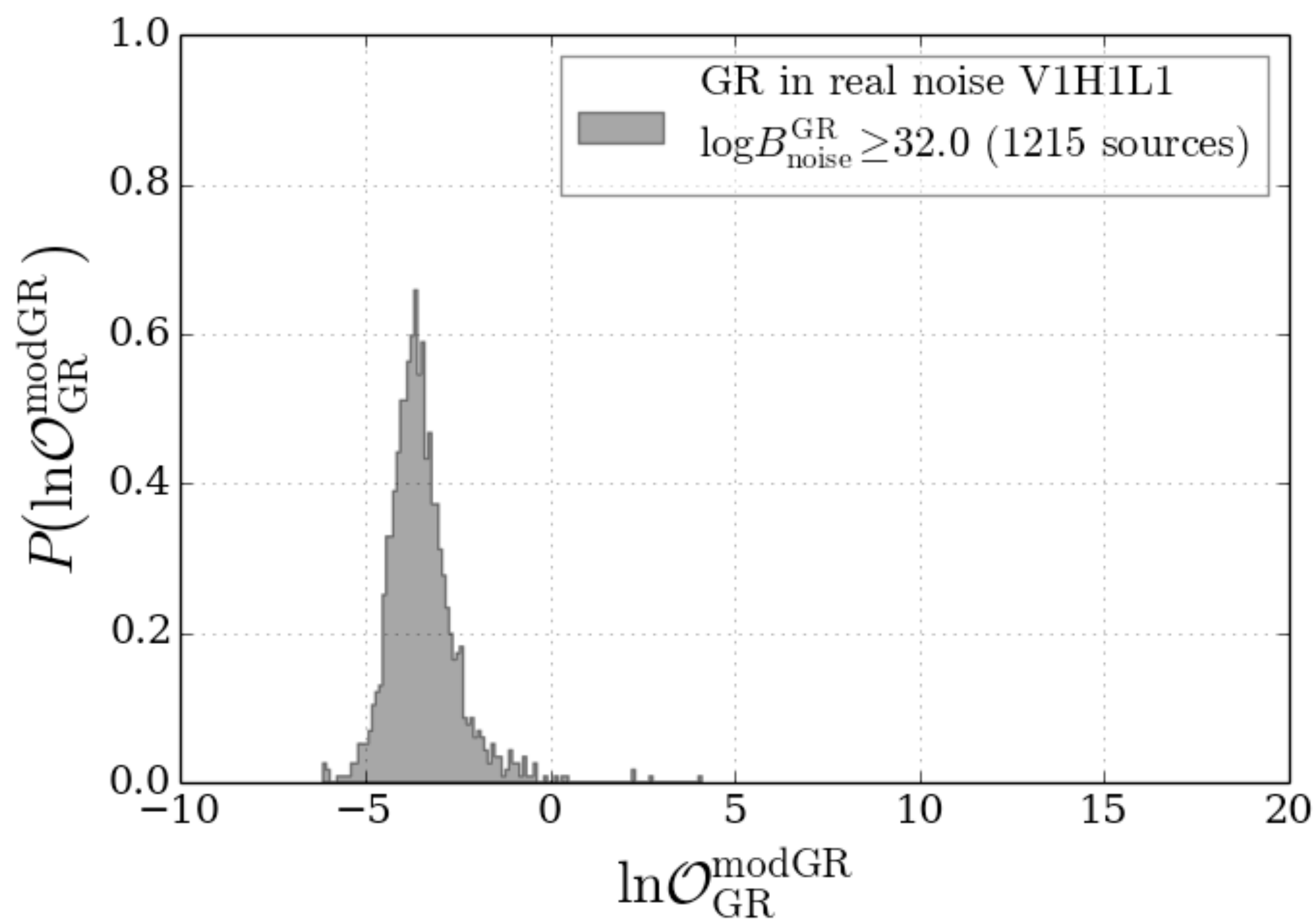




-10% shift at 1.5PN
catalogs of 15 sources

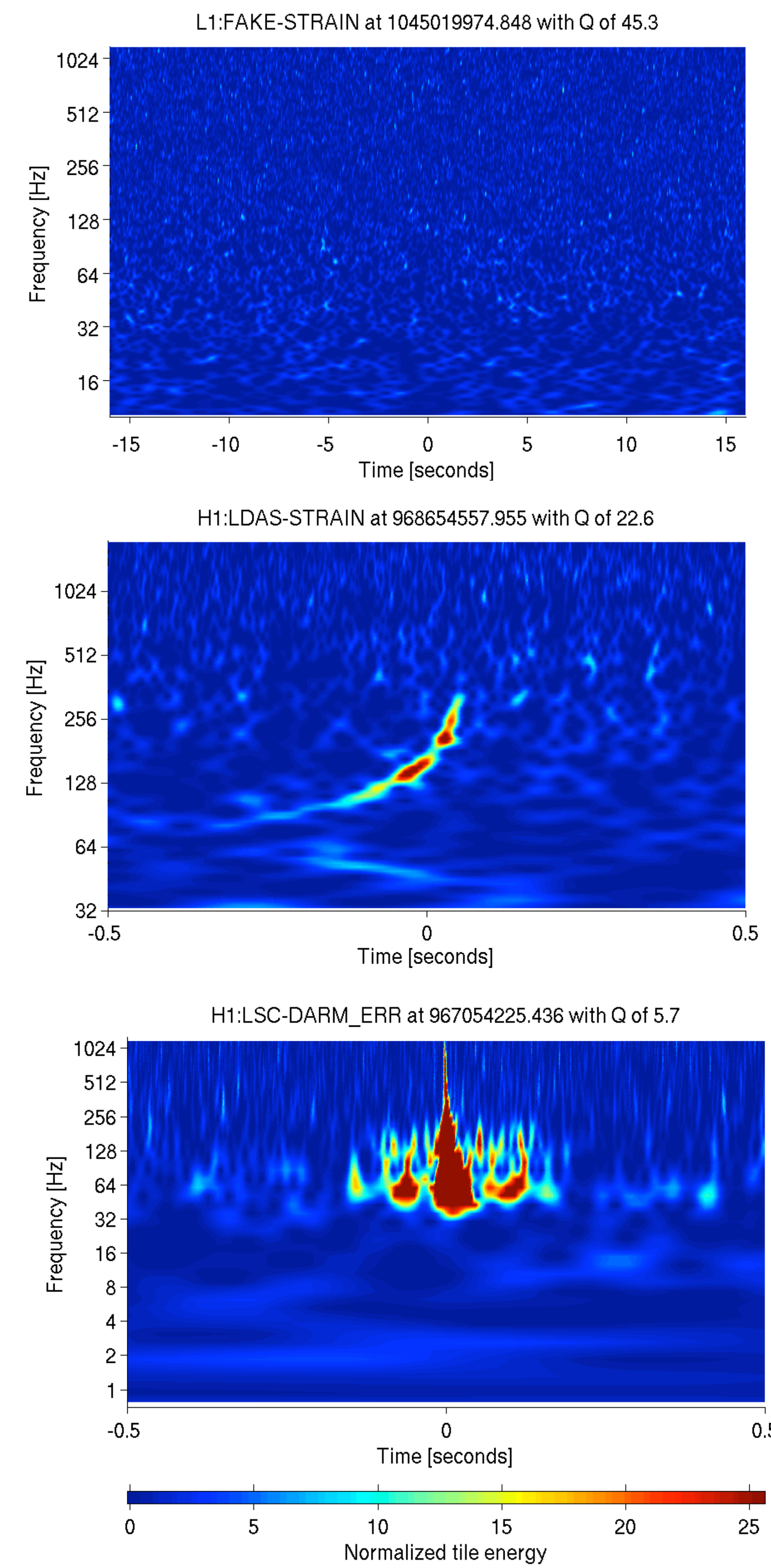
Realistic noise

- The above results used synthetic Gaussian noise based on design sensitivities
- In reality, noise will be **non-Gaussian** (glitchy) and **non-stationary** (varying PSD). Filter data by applying “vetoes”.
- Used S6,VSR2/3 data, recolored to aLIGO/AdVirgo noise curves after filtering out glitchy data segments
- Estimate PSD “on the fly” in the vicinity of the detection
- Ready for **O1**



Realistic noise (cont.)

- Non-Gaussian noise features (glitches) may compromise our GR test
- Glitches can be caused by environmental factors or temporary misbehaviour of detector component
- Rely on data quality algorithms to **veto** bad data segments and exclude glitches from analysis
- Correlate candidate glitches with data from auxiliary channels (microphones, magnetic sensors, accelerometers, etc.) to identify source



Extending TIGER to BBH/NSBH (preliminary)

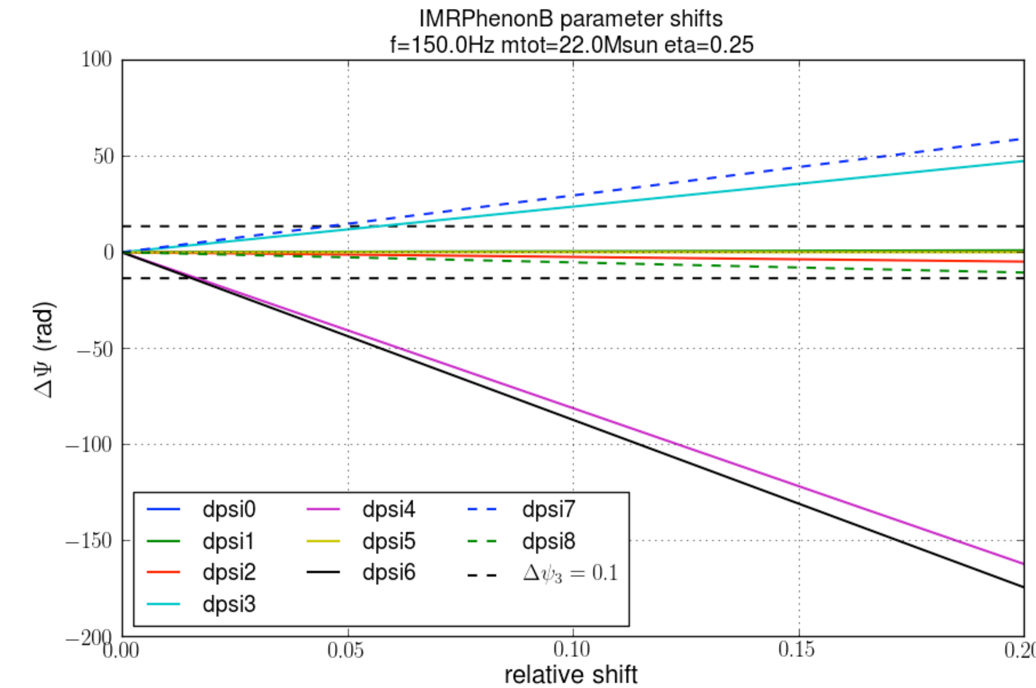
- need faithful waveforms that capture precession for background **injections** (e.g. *SEOBNR*, Pan et al. 2013 [arXiv:1307.6232])

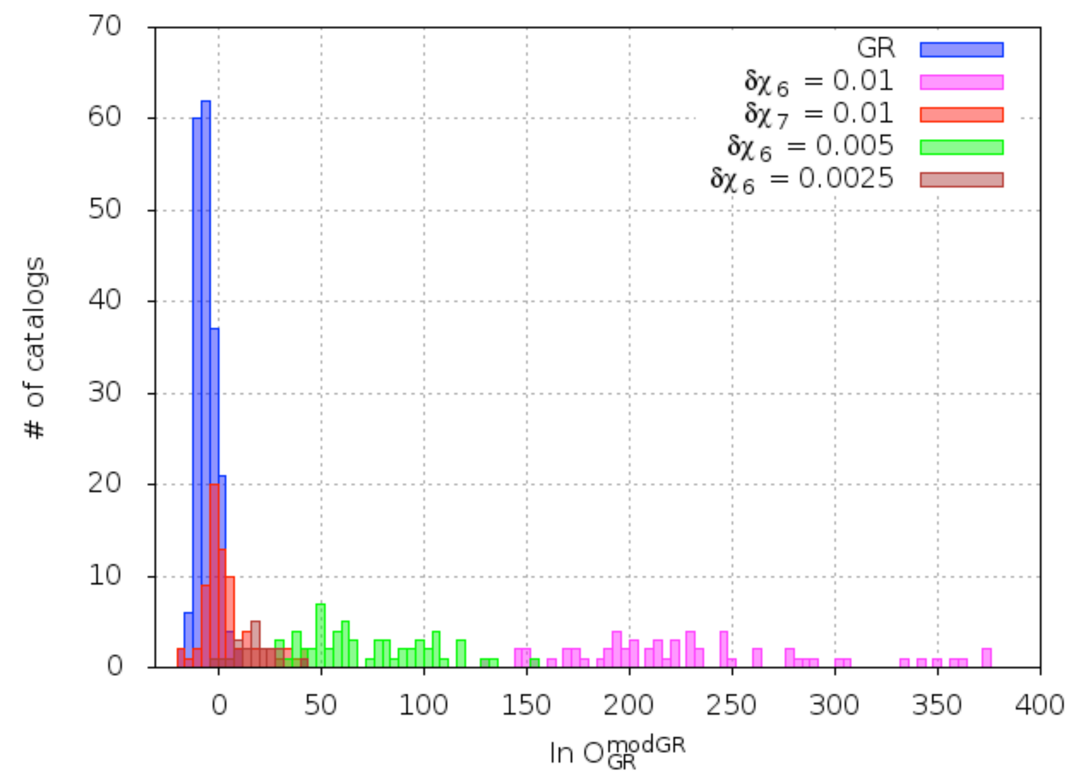
- need fast waveforms that capture precession for **recovery** (e.g. *IMRPhenomP* by Hannam et al. 2013 [arXiv:1308.3271])

- shorter waveforms but richer parameter space to be covered

- more susceptible to glitch contamination

- possibly use Reduced Order Modeling methods (Canizares et al. 2014 [arXiv:1404.6284])





What comes next?

- Possible outcomes for observed set of sources
  - GR seems to be favoured: start placing bounds on modGR parameters
  - marginal: wait for more data to arrive
  - GR seems to be violated: follow up with tailor-made tests to pinpoint the nature of violation; model selection against alternative theories, parameter estimation on their parameter space

- Alternative theories come with their own extended parameter space:
  - choose appropriate parameterization
  - identify locus not excluded by current bounds
  - identify locus that is distinguishable from GR
  - identify locus that does not modify signal "too much"

- Interesting scenarios: theories that predict no effect at weak/non-relativistic regime but may be observeable towards compact binary coalescence (phase transition *e.g.* spontaneous scalarization)

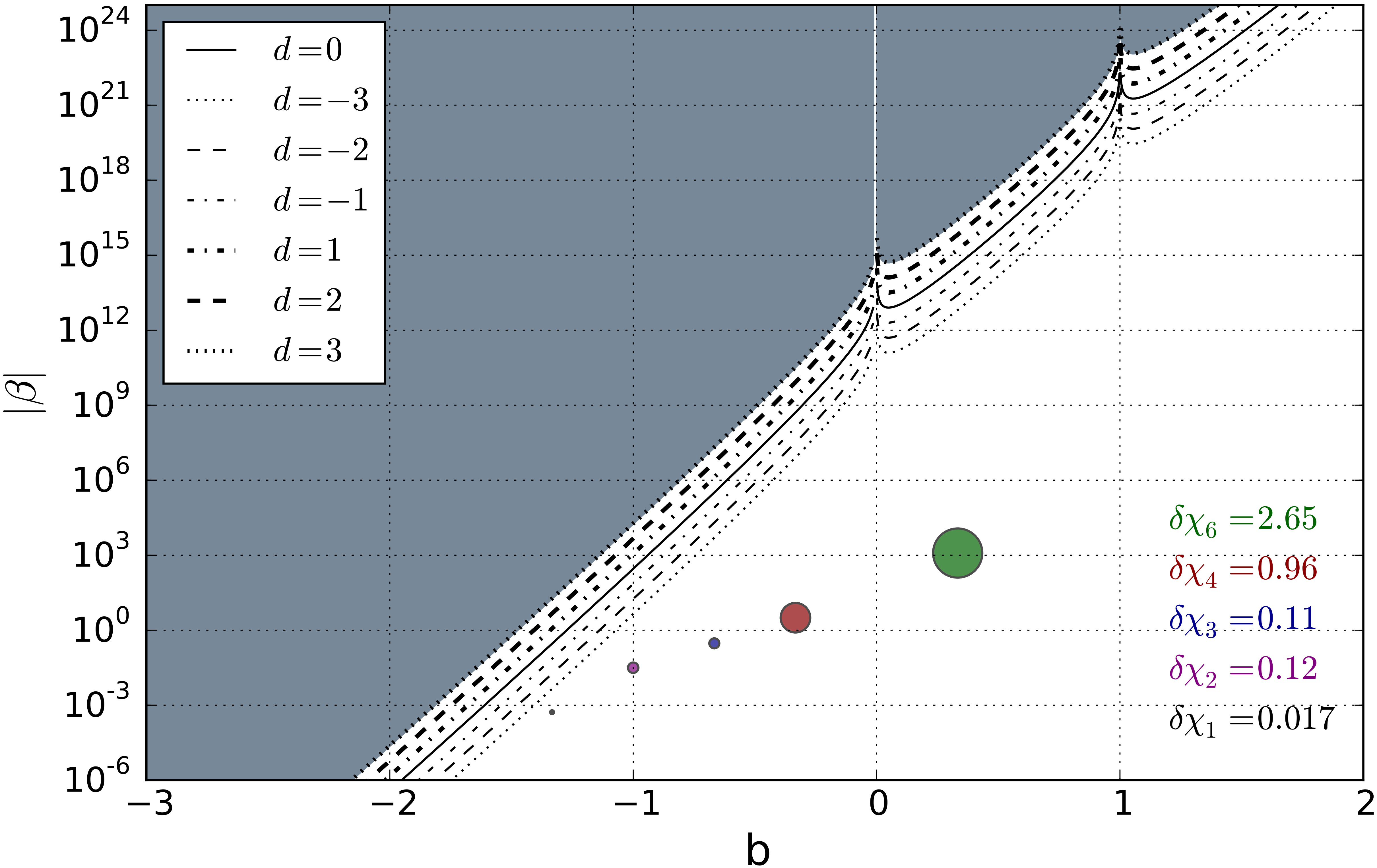
Conclusions

- TIGER is a Bayesian post-detection pipeline for BNS that **accumulates evidence** for finding possible violations of GR with unprecedented precision

- In the case of an observed GR-violation, more specific follow-up tests can be performed to narrow down the nature of the deviation (model selection & parameter estimation)

- Many different possible concerns have been addressed and the pipeline is shown to be robust

- extending TIGER to BBH/NSBH:
  - need faithful waveforms for background **injections**
  - need fast waveforms for **recovery**
  - Reduced Order Modeling methods



$d = 0$
$d = -3$
$d = -2$
$d = -1$
$d = 1$
$d = 2$
$d = 3$
$\delta\chi_6 = 2.65$
$\delta\chi_4 = 0.96$
$\delta\chi_3 = 0.11$
$\delta\chi_2 = 0.12$
$\delta\chi_1 = 0.017$
$|\beta|$
b
10^{24}
10^{21}
10^{18}
10^{15}
10^{12}
10^{9}
10^{6}
10^{3}
10^{0}
10^{-3}
10^{-6}
−3
−2
−1
0
1
2