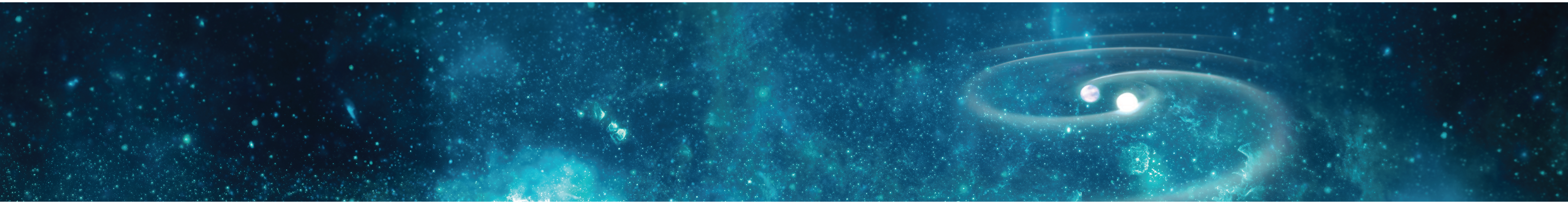


Data Analysis Challenges with LISA: Current status and prospects

N Karnesis
APC

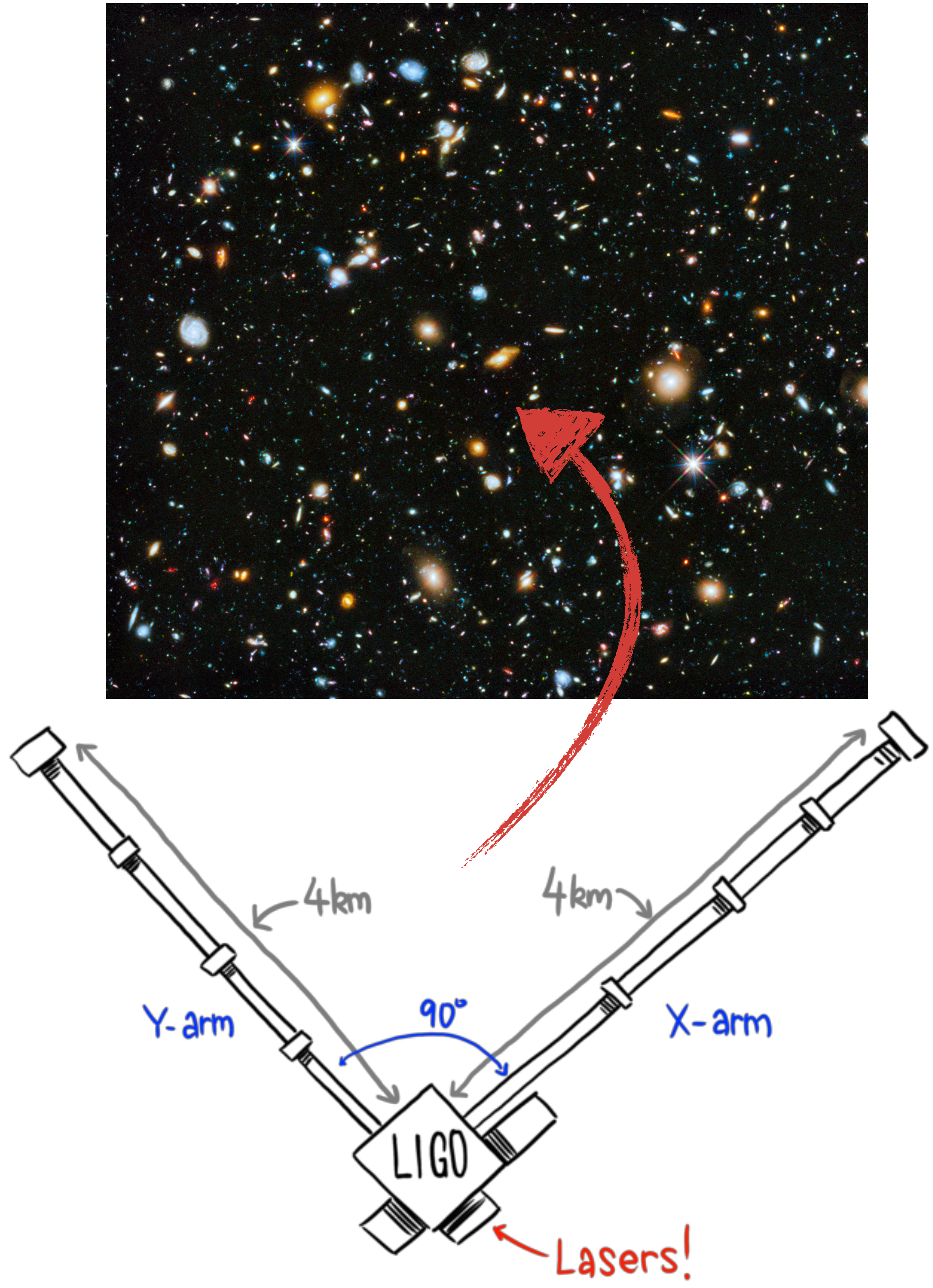
Rencontres GdR OG/ISIS
08-10-2018
Paris

Outlook

- LISA
- Data Analysis
  - Differences to the ground based detectors
  - Past, present & future

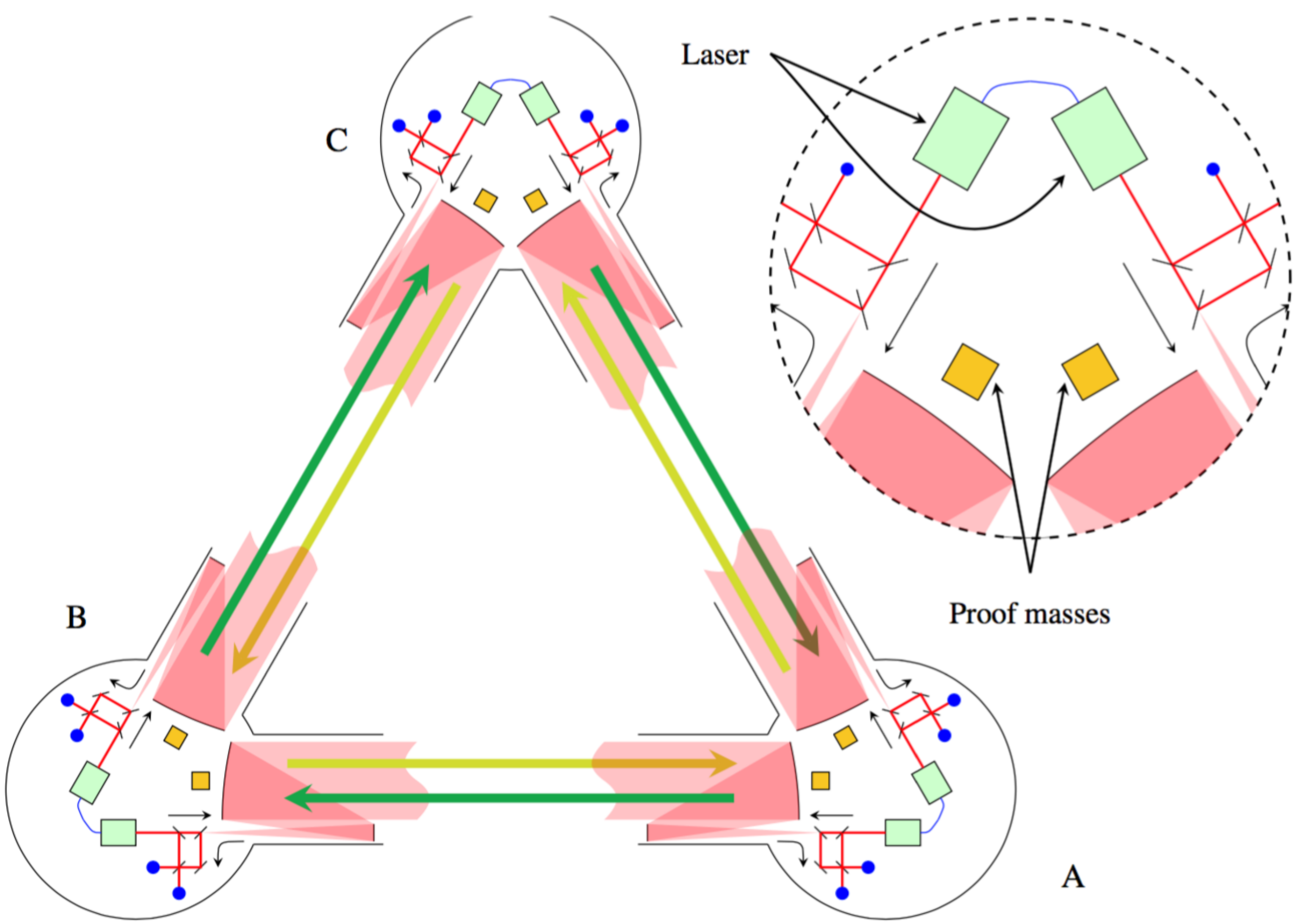
Laser Interferometer Space Antenna

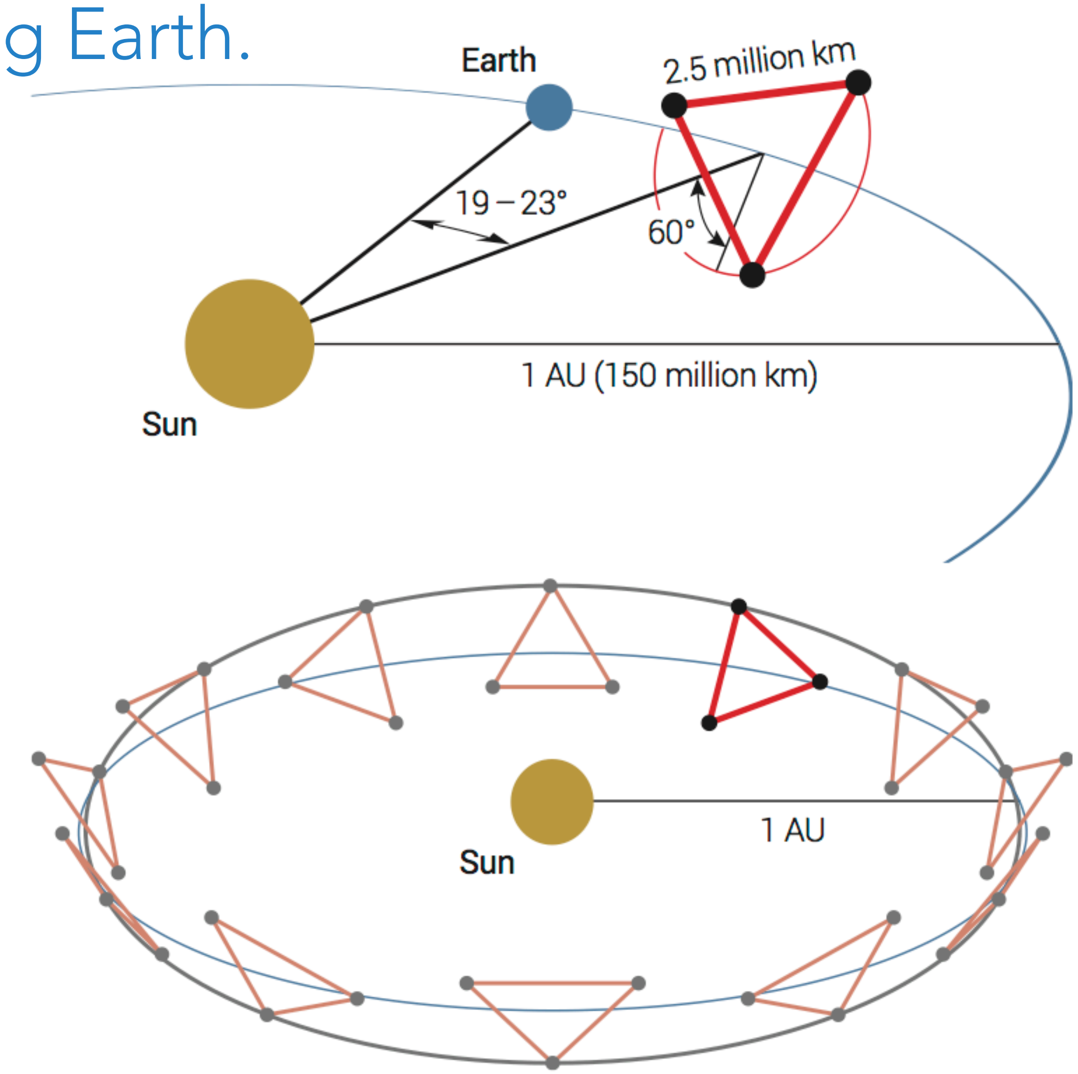
- LISA
  - The most simple way to look at the idea of LISA, is take the concept of a ground based detector, and put it into space.
  - But this is a *very* rough approximation. LISA is very different from GBD in the
    - Nature of the data stream
    - Science Operations
    - Types of Sources
    - Data Analysis requirements



Laser Interferometer Space Antenna

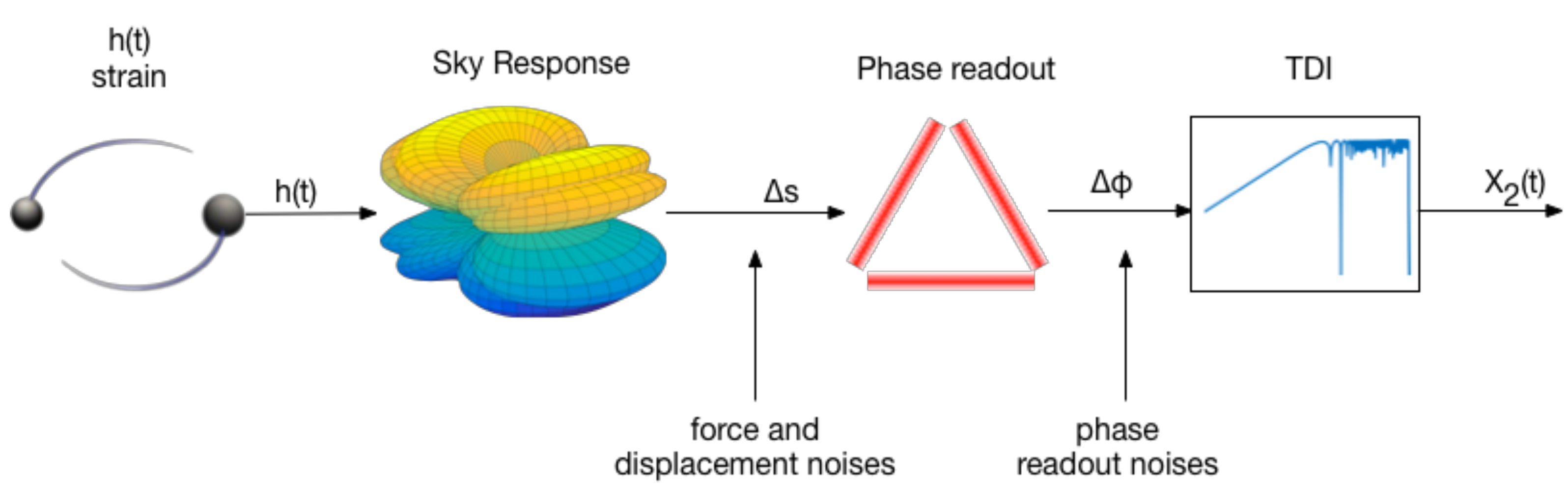
- Nature of the data stream/Science Operations:
  - A constellation of three Space Crafts, forming an equilateral triangle with arms of 2.5^6 km.
  - Each SC contains free falling test masses where their position is being constantly monitored via means of laser interferometry.
  - The constellation orbits the sun following Earth.





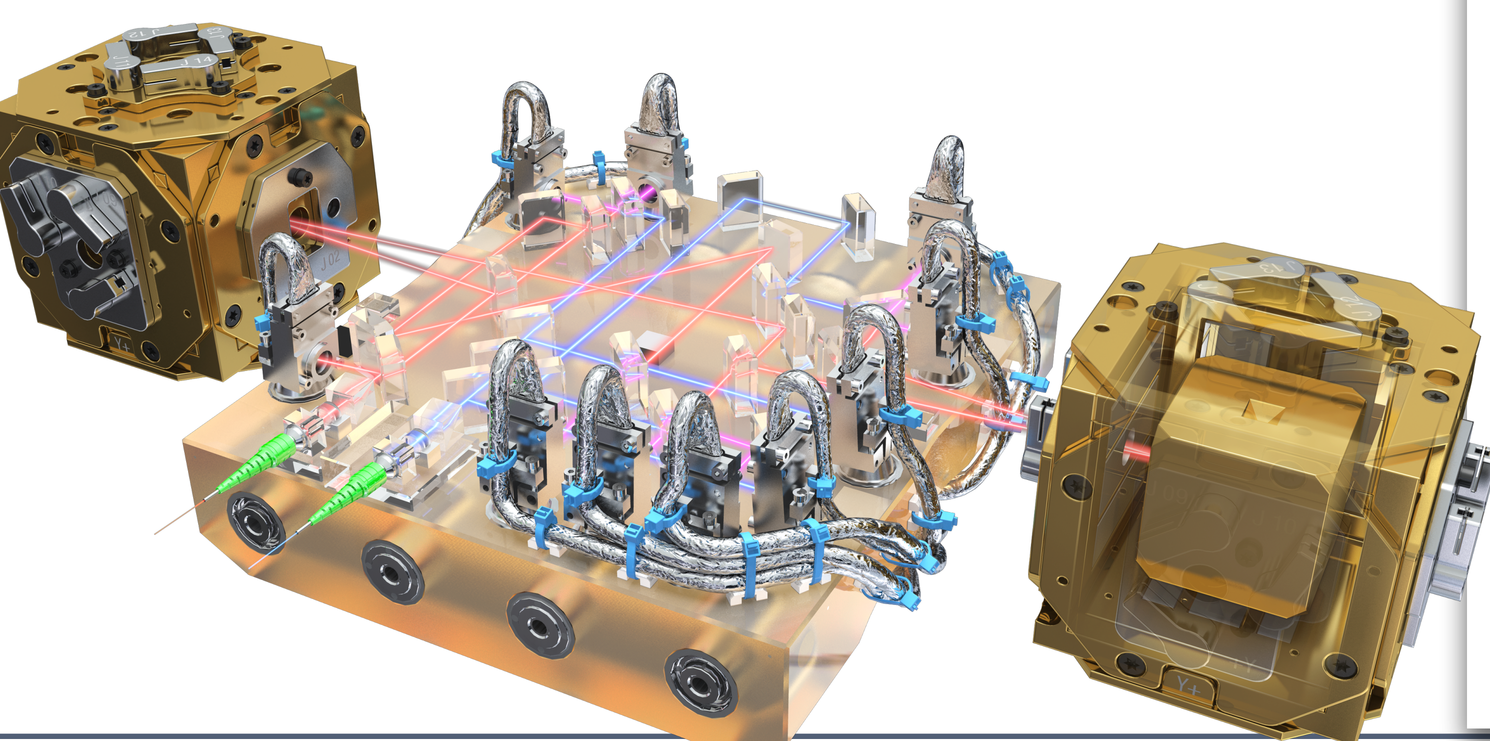
Laser Interferometer Space Antenna

- Nature of the data stream/Science Operations:
  - Data size is quite small (low rate)
  - Before SCi Op we need to point, calibrate, optimise (~1 yr of commissioning).
  - We need to perform a Time Delayed Interferometry (TDI) algorithm to get rid of the laser frequency noise.
  - Ideally undisturbed measurements for timespans of 10 days, allowing for short ~hr data gaps for SC maintenance.
  - Data artefacts (lines, glitches, bursts)
- All the above sound good, but how could be proven that LISA is going to be successful?

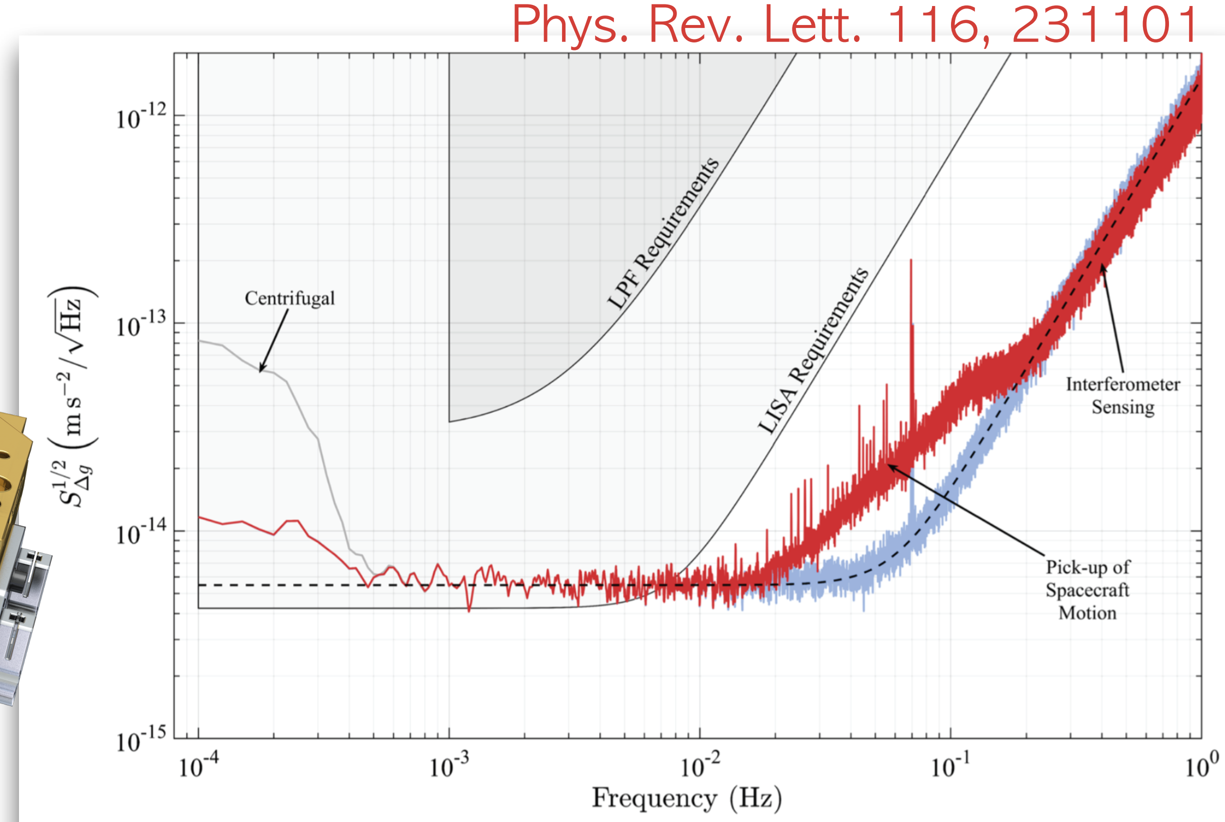


LISA Pathfinder (LPF)

- But how could be proven that LISA was going to be successful?
  - ESA decided to fly the LPF mission, a technology demonstrator for LISA (Launched Dec 2015, decommissioned Jul 2017).
  - Basically a lab in space, shrinks a LISA arm from 2.5^6 km to 30 cm.
  - Aim: prove technologies & characterise and model noise sources.

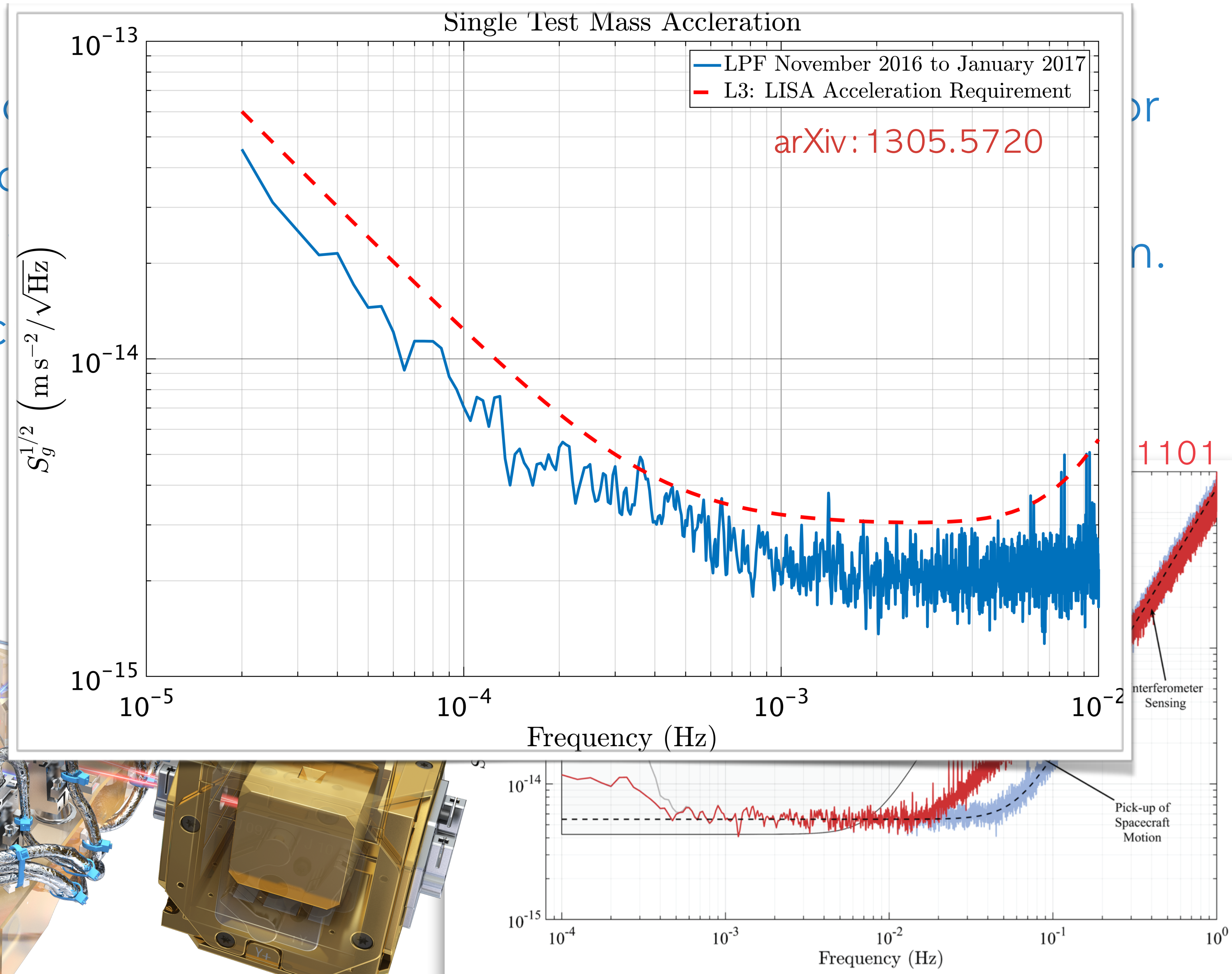


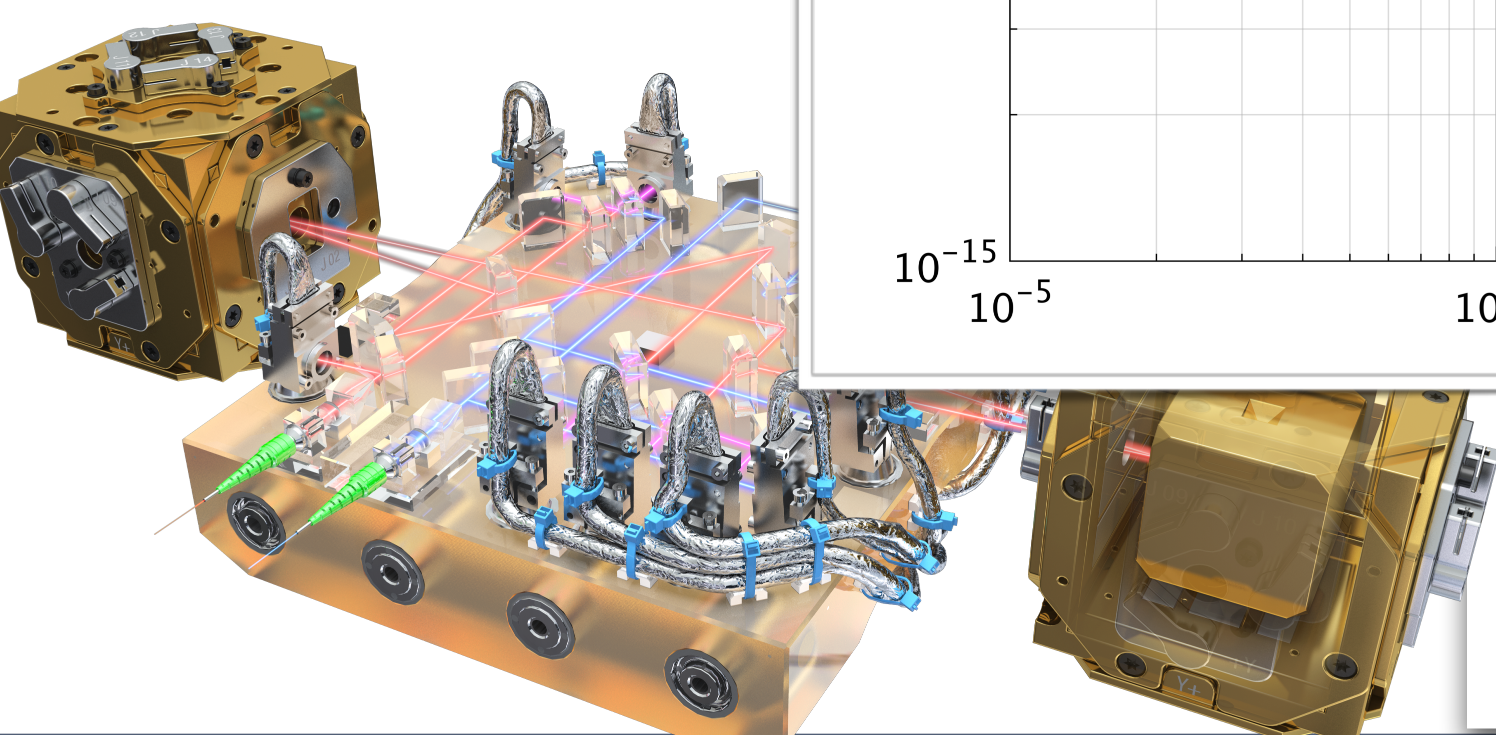
Phys. Rev. Lett. 116, 231101



LISA Pathfinder (LPF)

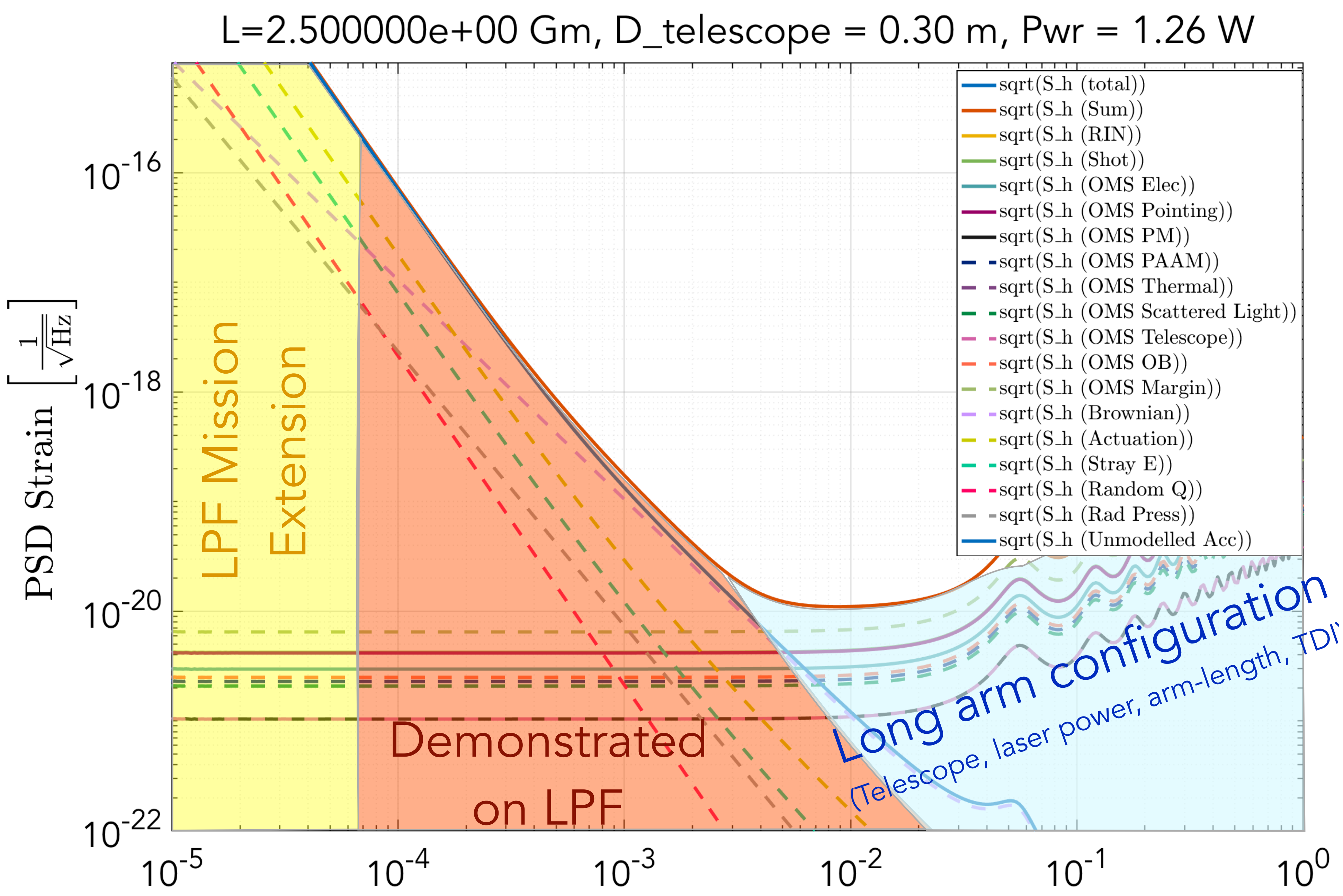
- But how could be proven that LISA was going to be successful?
  - ESA decided t… …or LISA (Launche…
  - Basically a lab … …n.
  - Aim: prove tec…





Laser Interferometer Space Antenna: noise

- Now we have a more complete prediction of the noise of the instrument.



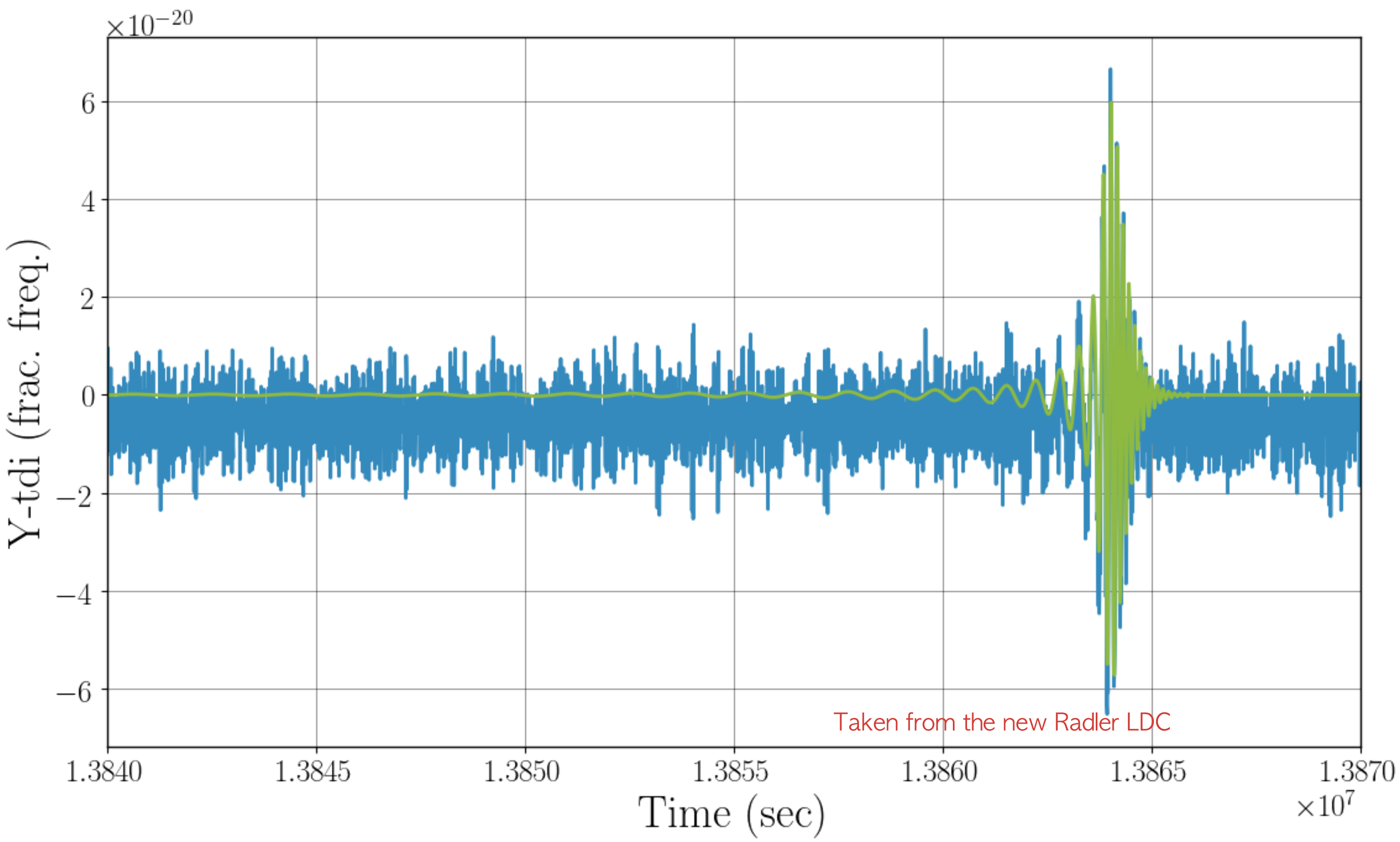
By M. Hewitson

Laser Interferometer Space Antenna: Sources

- And being more confident about the noise, we start to list the possible sources of the GW signals:
  - Massive Black Hole Binaries (MBHB)
  - Extreme Mass Ratio Inspirals (EMRIs)
  - Galactic Binaries
  - Stochastic GW Background

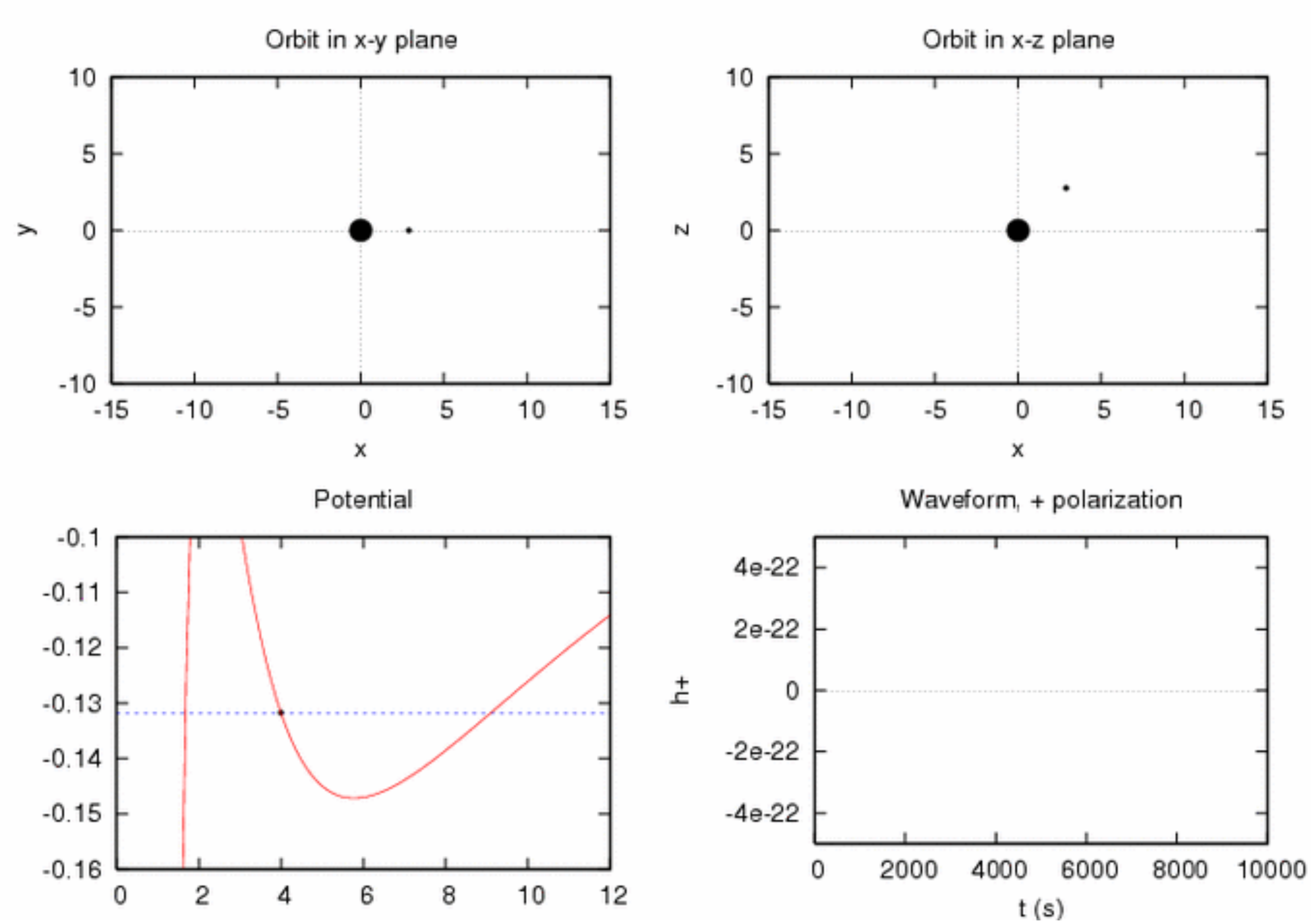
Sources: MBHB

- Expect events with high SNR
- Signal duration varies from weeks to ~year.
- Typically ~15 dimensions parameter space.



Sources: EMRIs

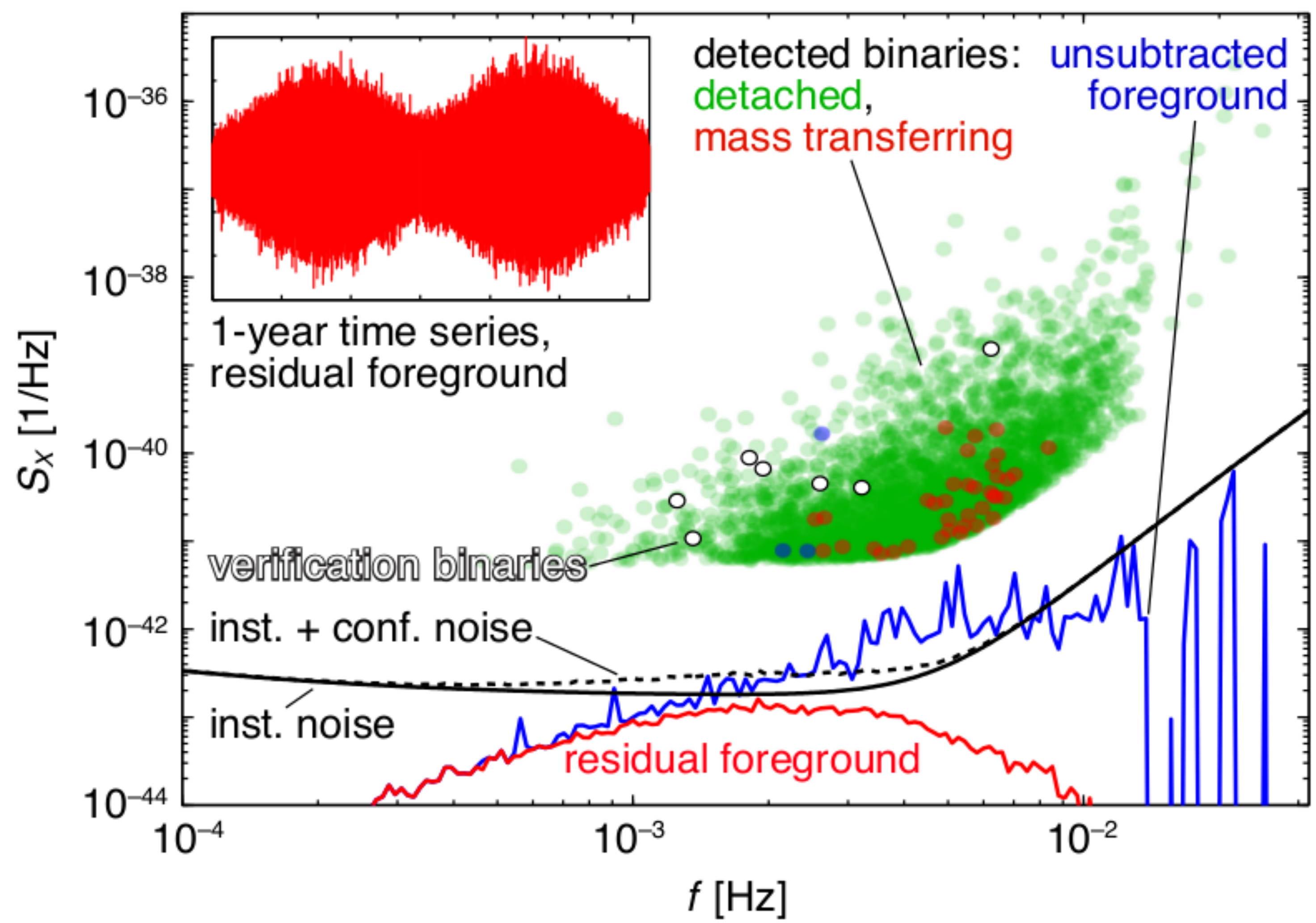
- Events with typical mass ratio of 10^-5 to 10^-7
- Usually a pair of MBH with a NS, BH, or WD.
- The result is a huge number of orbits in the LISA band and a characteristic waveform.



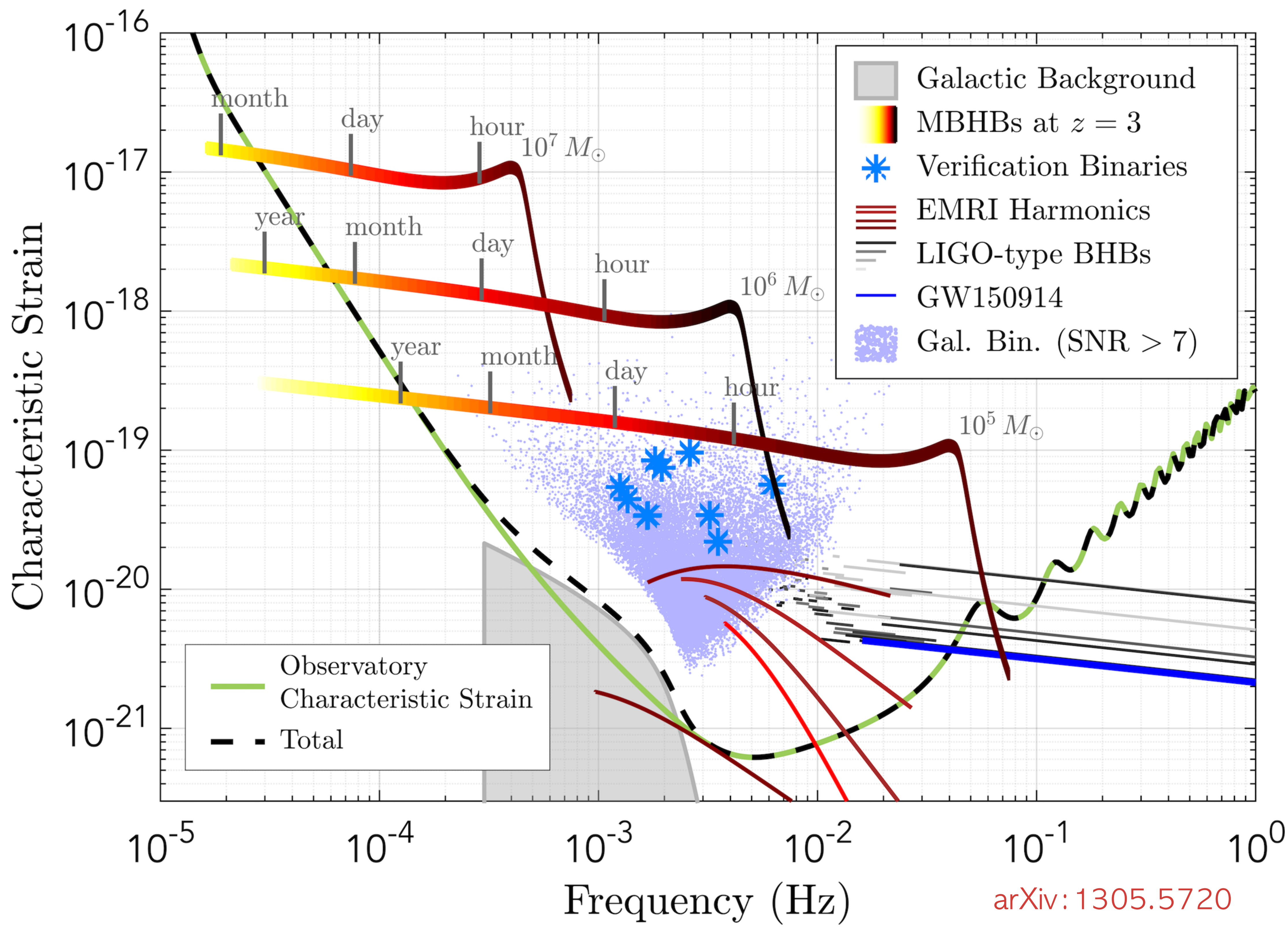
By J. Gair

Sources: Galactic White Dwarf Binaries

- Expect 10^7 binaries with slow evolution in frequency.
- Almost monochromatic
- Verification Binaries - calibration.



Laser Interferometer Space Antenna: Sources



arXiv:1305.5720

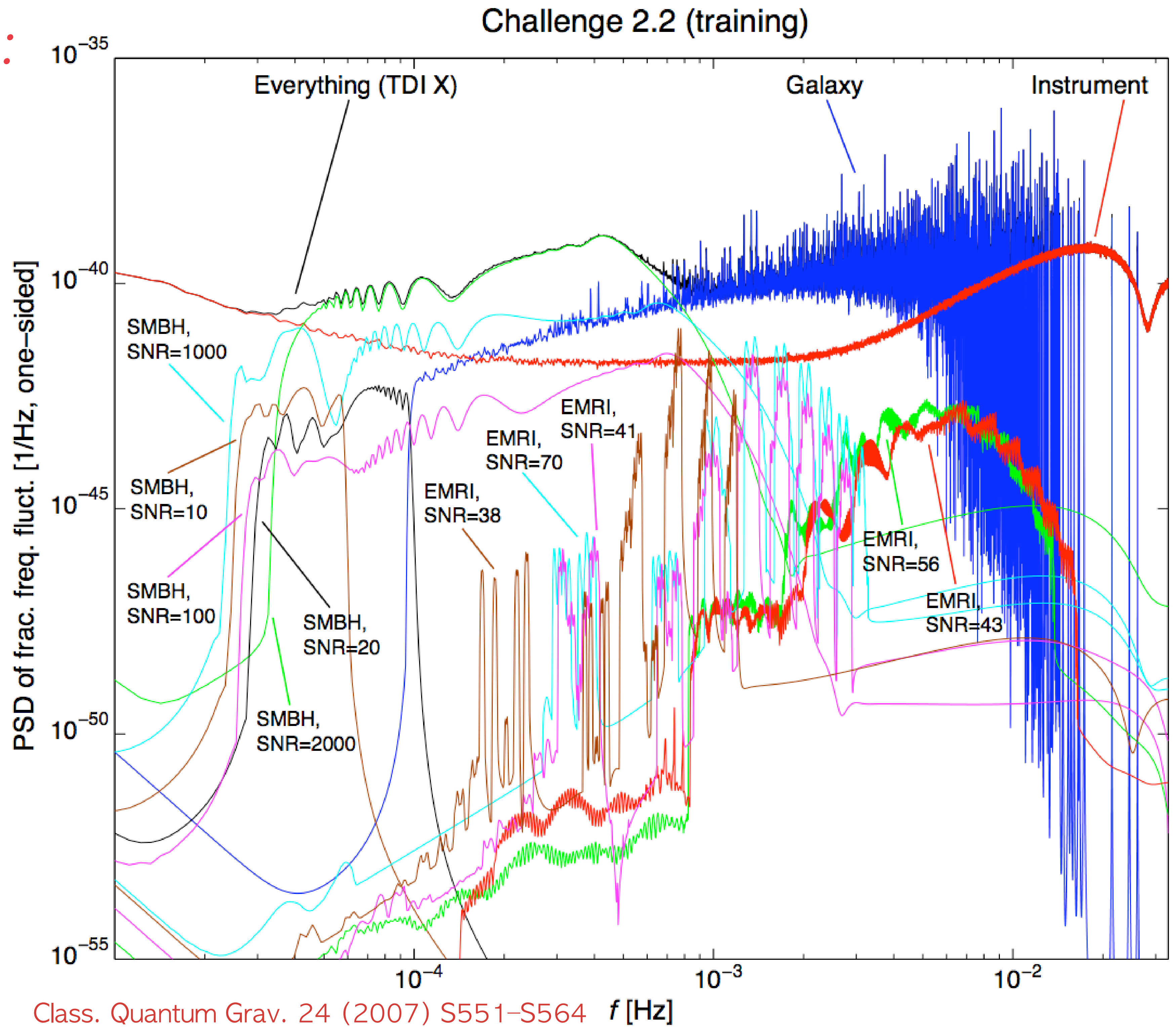
Complete Signal

- To sum them up, we expect
  - MBHB: 10 -100 per yr
  - EMRIs: 10 -1000 per yr
  - SOBHB: 30 - 100 per yr
  - GB: ~60 millions

Complete Signal

- Now we can sum up all the predicted sources in what we call the *enchilada:*

Source: http://tsgcookin.com



Class. Quantum Grav. 24 (2007) S551–S564

Data Analysis

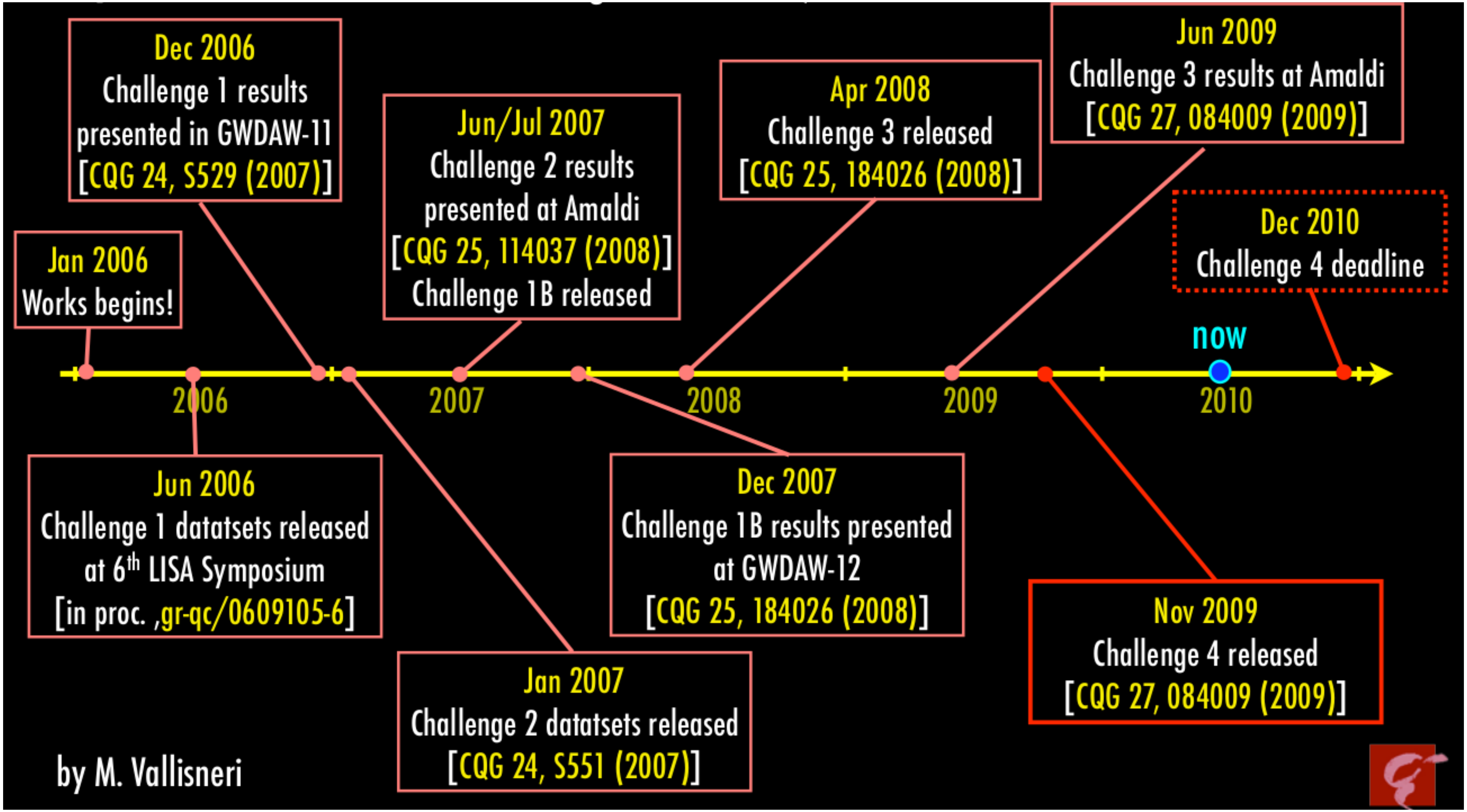
- For that reason, the LISA Data Challenges (LDCs) have been organised.
- Solve the "enchilada", also take into account the parametrisation of the noise.
  - Which causes further increase of the dimensionality of the problem.
- Matched filtering has been widely used for the analysis.
  - Model the sources and search the data stream for patterns.
  - Need to "whiten" the data
    - Calibrate to the detectors' noise spectral shape.
- Make use of Bayes Theorem:

$$p(\vec{\theta}|d) = \frac{p(d|\vec{\theta}, \mathcal{M})p(\vec{\theta})}{p(d|\mathcal{M})}.$$

- Necessary tools need to be adopted / developed.
  - So far, stochastic search methods are preferred to the more straightforward grid methods (huge parameter space).
    - Traditional MCMC and its variants, PT, Nested Sampling (MultiNest, CPNest, PolyChord , …), Multimodal Genetic algorithm.

LISA Data Challenges - The Past

- Started 2006, last one at 2010.

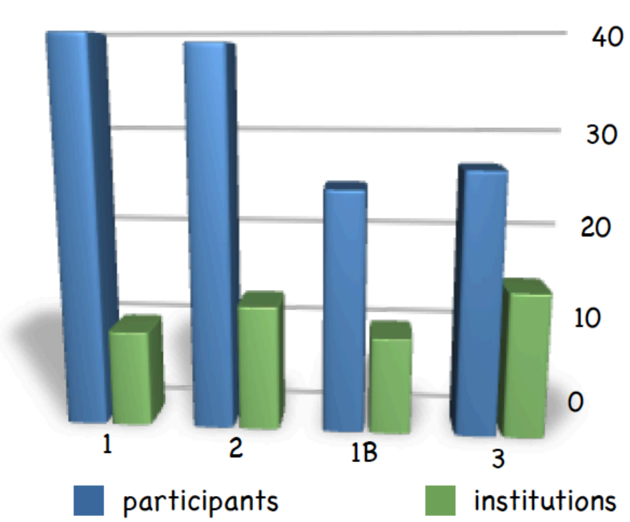


LISA Data Challenges - The Past

- The attendees:

70 Participants from 25 institutions

- Albert Einstein Institute Golm
- Albert Einstein Institute Hannover
- APC, CNRS, Paris
- ARTEMIS, CNRS, Nice
- U. of Auckland
- U. of Birmingham
- Caltech/NASA JPL
- U. of Cambridge
- Cardiff U.
- Carleton College
- Chinese Academy of Science, Bejing
- U. of Glasgow
- NASA Ames
- NASA Goddard
- U. Iles Balears
- Indian Inst. of Tech., Kharagpur
- U. Maryland
- Montana State U.
- Nanjing U.
- Northwestern U.
- Polish Academy of Science
- Rochester Institute of Technology
- U. of Texas Brownsville
- U. of Southampton



	participants	institutions
MLDC 1	40	10
MLDC 2	39	13
MLDC 1B	25	10
MLDC 3	27	15

By S. Babak

LISA Data Challenges - The Past

- The menu:

	MLDC 1	MLDC 2	MLDC 1B	MLDC 3	MLDC 4
Galactic binaries	• Verification • Unknown isolated • Unknown interfering	• Galaxy $3x10^6$	• Verification • Unknown isolated • Unknown interfering	• Galaxy $6x10^7$ chirping	• Galaxy $6x10^7$ chirping
Massive BH binaries	• Isolated	• 4-6x, over "Galaxy" & EMRIs	• Isolated	• 4-6x spinning & precessing over "Galaxy"	• 4-6x spinning & precessing, extended to low-mass
EMRI		• Isolated • 4-6x, over "Galaxy" & MBHs	• Isolated	• 5 together, weaker	• 3 x Poisson(2)
Bursts				• Cosmic string cusp	• Poisson(20) cosmic string cusp
Stochastic background				• Isotropic	• Isotropic

LISA Data Challenges - The Past

- The results for MBHB as an example:

source ($\mathrm{SNR_{true}}$)	group	$\delta M_c/M_c$ $\times 10^{-5}$	$\delta\eta/\eta$ $\times 10^{-4}$	δt_c (sec)	δsky (deg)	δa_1 $\times 10^{-3}$	δa_2 $\times 10^{-3}$	$\delta D/D$ $\times 10^{-2}$	FF
MBH-1	AEI	2.4	6.1	62.9	11.6	7.6	47.4	8.0	0.9936
(1670.58)	CambAEI	3.4	40.7	24.8	2.0	8.5	79.6	0.7	0.9925
	MTAPC	24.8	41.2	619.2	171.0	13.3	28.7	4.0	0.9996
	JPL	40.5	186.6	23.0	26.9	39.4	66.1	6.9	0.9981
	GSFC	1904.0	593.2	183.9	82.5	5.7	124.3	94.9	0.1827
MBH-3	AEI	9.0	5.2	100.8	175.9	6.2	18.6	2.7	0.9995
(847.61)	CambAEI	13.5	57.4	138.9	179.0	21.3	7.2	1.5	0.9993
	MTAPC	333.0	234.1	615.7	80.2	71.6	177.2	16.1	0.9945
	JPL	153.0	51.4	356.8	11.2	187.7	414.9	2.7	0.9898
	GSFC	8168.4	2489.9	3276.9	77.9	316.3	69.9	95.6	0.2815
MBH-4	AEI	4.5	75.2	31.4	0.1	47.1	173.6	9.1	0.9994
(160.05)	CambAEI	3.2	171.9	30.7	0.2	52.9	346.1	21.6	0.9991
	MTAPC	48.6	2861.0	5.8	7.3	33.1	321.1	33.0	0.9352
	JPL	302.6	262.0	289.3	4.0	47.6	184.5	28.3	0.9925
	GSFC	831.3	1589.2	1597.6	94.4	59.8	566.7	95.4	−0.1725
MBH-2	AEI	1114.1	952.2	38160.8	171.1	331.7	409.0	15.3	0.9469
(18.95)	CambAEI	88.7	386.6	6139.7	172.4	210.8	130.7	24.4	0.9697
	MTAPC	128.6	45.8	16612.0	8.9	321.4	242.4	13.1	0.9260
	JPL	287.0	597.7	11015.7	11.8	375.3	146.3	9.9	0.9709
MBH-6	AEI	1042.3	1235.6	82343.2	2.1	258.2	191.6	26.0	0.9293
(12.82)	CambAEI	5253.2	1598.8	953108.0	158.3	350.8	215.4	29.4	0.4399
	MTAPC	56608.7	296.7	180458.8	119.7	369.2	297.6	25.1	0.0016

Babak et al, arXiv:0912.0548

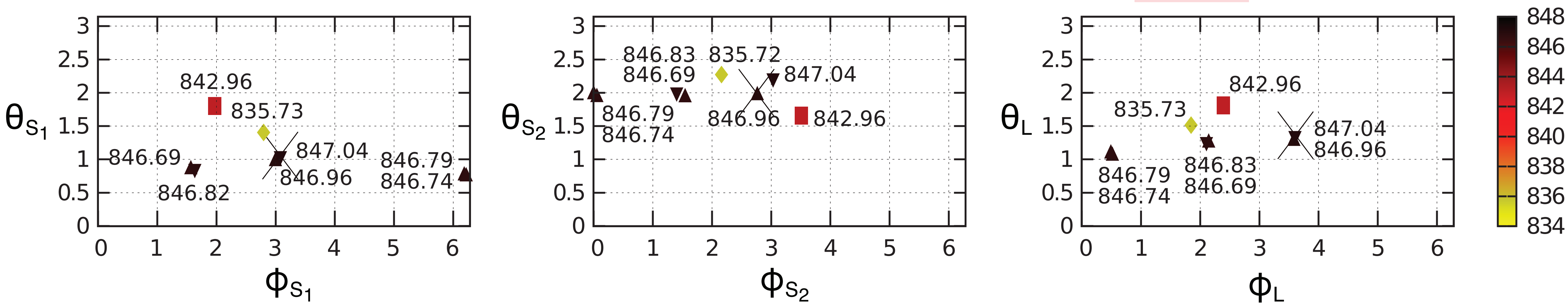
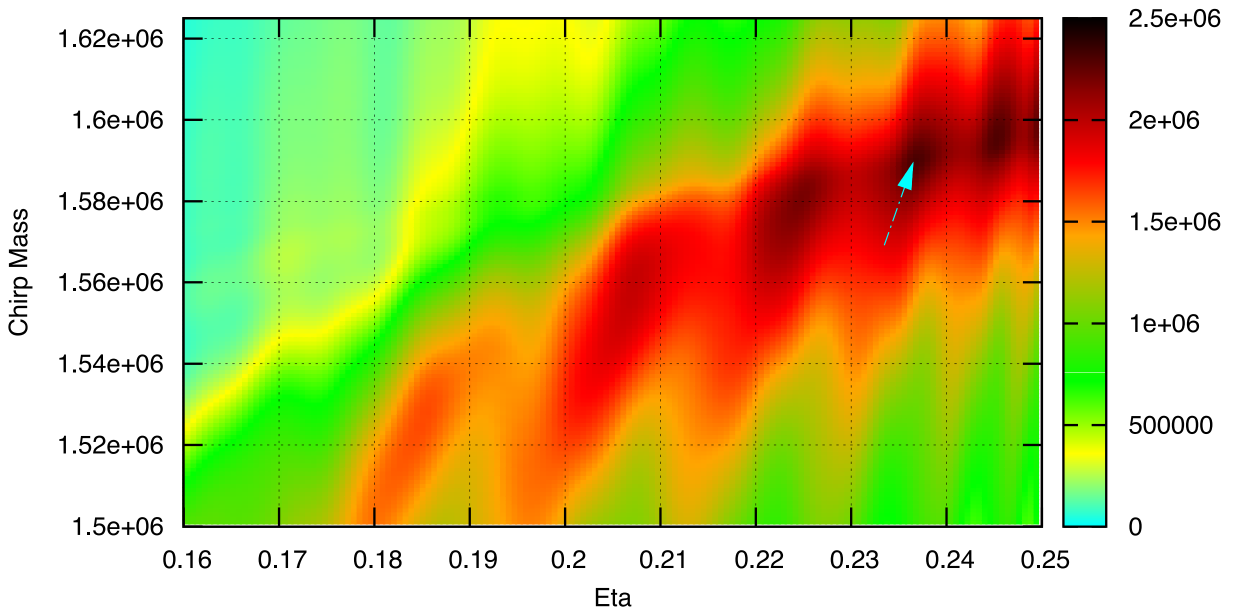
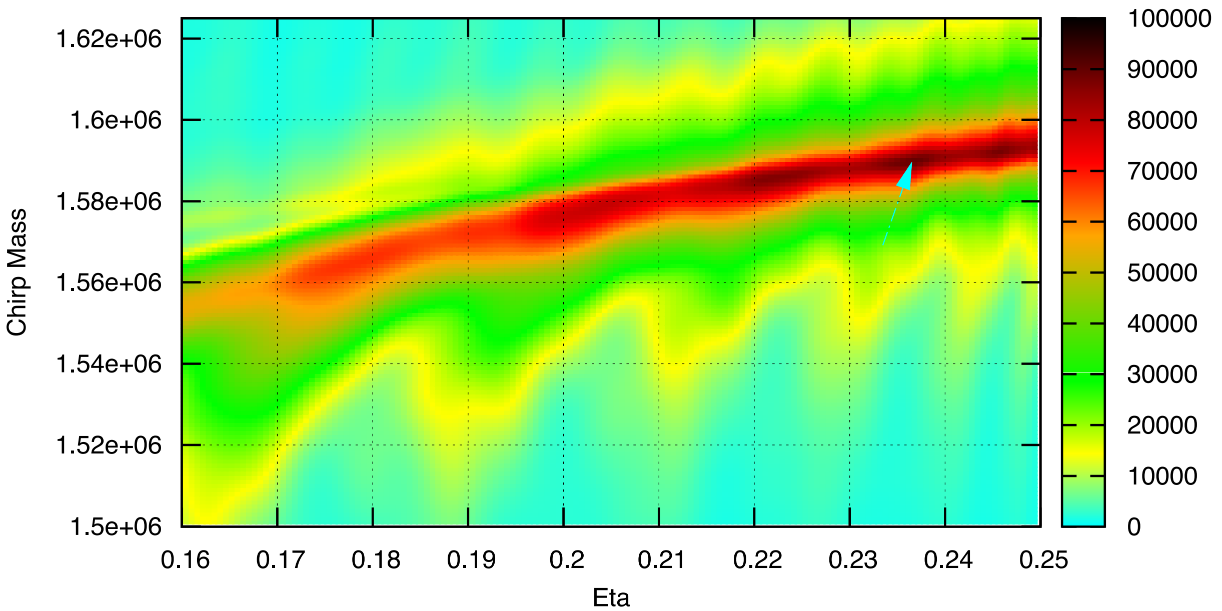


Figure 1. Distribution of reported modes for MBH-3 along the spin and orbital–angular-momentum angles. Each mode is annotated with the SNR. ×: true value;

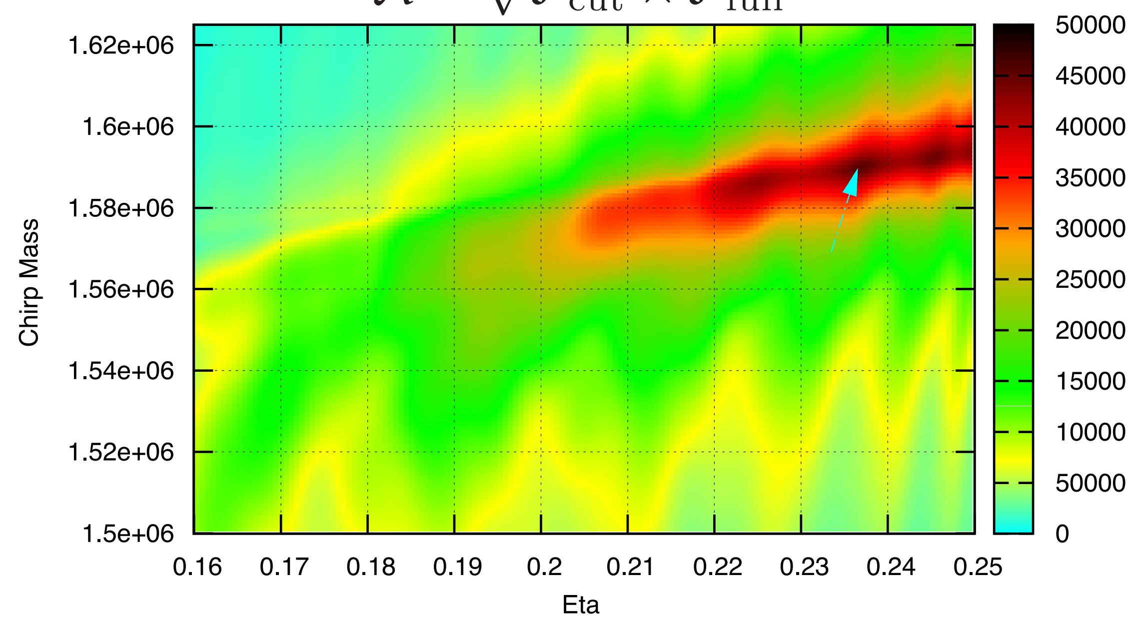
LISA Data Challenges - The Past

- The results for MBHB:





$$\mathcal{A} = \sqrt{\mathcal{F}_{\mathrm{cut}} \times \mathcal{F}_{\mathrm{full}}}$$



Petiteau et al, PRD 81, 104016, 2010

LDCs and What have we learned so far?

- Multimodal posterior surfaces.
- High dimensionality.
- Source separation.

→ Perform **Model Selection** at the same time.

- Efficient samplers/methods.
  - Efficiently map the posterior surface.
- Take into account non-stationarity of the noise.
  - Parametrisation of slow fluctuation, bursts, transients.
- Iterative Global Fit Scheme:
  - As data is transmitted on ground, we update uncertainties, discover new signals, improve the quality of the fit.

LISA Data Challenges - The Present

LISA Data Challenges - The Present

- The last couple of years things have started to accelerate again.
  - GW Detection & events, LPF, LISA L3 selection
- The LDC Working Group
  - Part of the Consortium.
- The "*Radler*" data sets are already online:

- Aim is to gather the people again, and get up to speed.
- Resurrect & update old ideas, propose & develop new ideas, gather the tools and codes for the analysis.
- Re-start the LDC in a new framework under the wings of the Consortium.
- Establish a common playground to test algorithms
- Introduce software standards

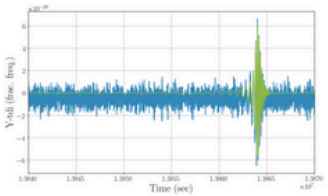
LISA Data Challenges - The Present

- The "*Radler*" data sets are already online:

https://lisa-ldc.lal.in2p3.fr/ldc

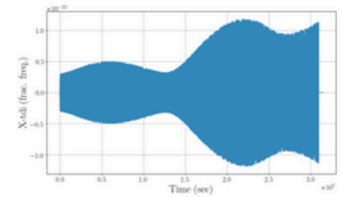
LDC1-1. A single GW signal from a merging massive-black-hole binary.

LIGO and Virgo have done it, so let's get LISA on the right path! MBHBs are represented with a frequency-domain inspiral–merger–ringdown phenomenological model (IMRPhenomD). The black holes are spinning, with spin vectors parallel to the orbital angular momentum. The release includes datasets for two methods (frequency- and time-domain) of applying the LISA response to the GWs.



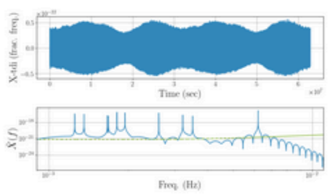
LDC1-2. A single GW signal from an extreme-mass-ratio inspiral.

EMRIs are modeled with the "classic" *Analytic Kludge* waveforms, which will be updated in future challenges, so make your code flexible! The signal is produced in the time domain and the response is applied using LISACode. The signal is of moderate strength, but the source parameters are drawn from relatively wide priors. This should make for a good challenge!



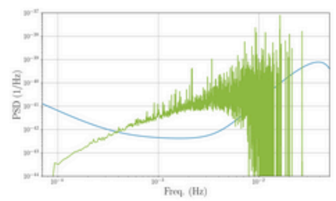
LDC1-3. Superimposed GW signals from several verification Galactic white-dwarf binaries.

We assume circular orbits and purely gravitational interactions. The phase of the signal includes frequency and first derivative. This one should be easy!



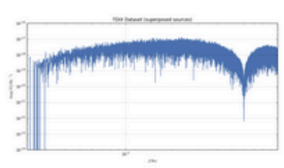
LDC1-4. A GW signal from a population of Galactic white-dwarf binaries.

Here's the classic cocktail-party problem: 26 million signals, produced with a "fast response" code. Parameters of all binaries are available in a large HDF5 file.



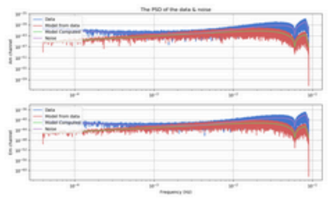
LDC1-5. A GW signal from a population of stellar-origin (stellar-mass) black-hole binaries.

LIGO and Virgo's gift to LISA. The population follows Salpeter's mass function, with an overall rate based on recent LIGO–VIRGO estimates. Waveform and LISA response are computed in the frequency domain.



LDC1-6. An isotropic stochastic GW signal of primordial origin.

Statistics are Gaussian, but the spectral shape is shrouded in mystery, with parameters chosen for us by the LISA Consortium Cosmology Working Group. The signal is generated using LISACode as a choir of elementary sources uniformly distributed across the sky. To make things easier for you, instrumental noise is Gaussian, uncorrelated, and of the same level in each LISA link.



- Join our efforts!
- LDC Code and docker images to download.
- Results are submitted &
- Tutorials & examples to be published soon.

The future

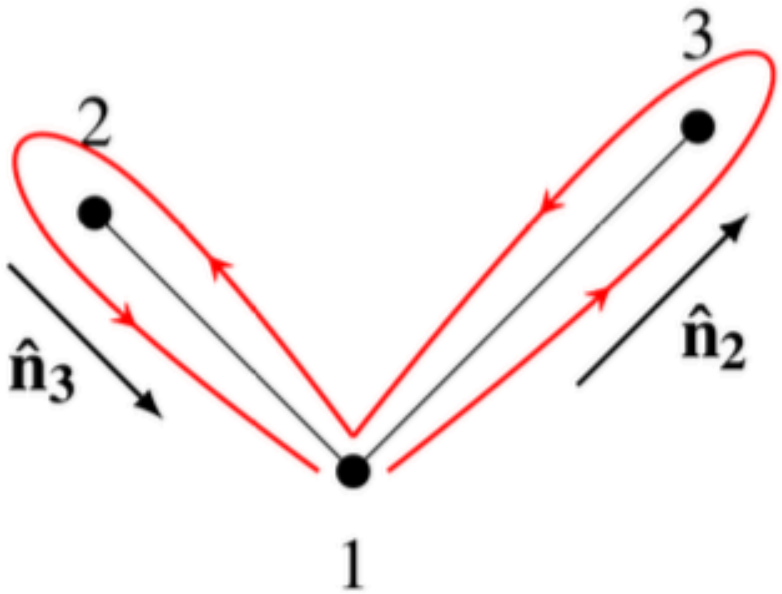
- LDC2: Increase realism with a "mild enchilada"
  - Mix of sources: Galaxy + MBHB + EMRIs, Galaxy + Stochastic + SOBHBs
- More realistic scenario from the point of instrument characteristics
- Release source catalogues
- Support further development activities.

Summary

- Signal from the LISA instrument is of different nature compared to the ground based detectors case.
  - Signal dominated data, long lived overlapping signals.
  - High dimensionality, multimodality
    - source separation & model selection.
- Work has been done within the LDC context.
- Now in the process of reviving the community
  - And the tools and methods that come together.
- Welcome to join our efforts!

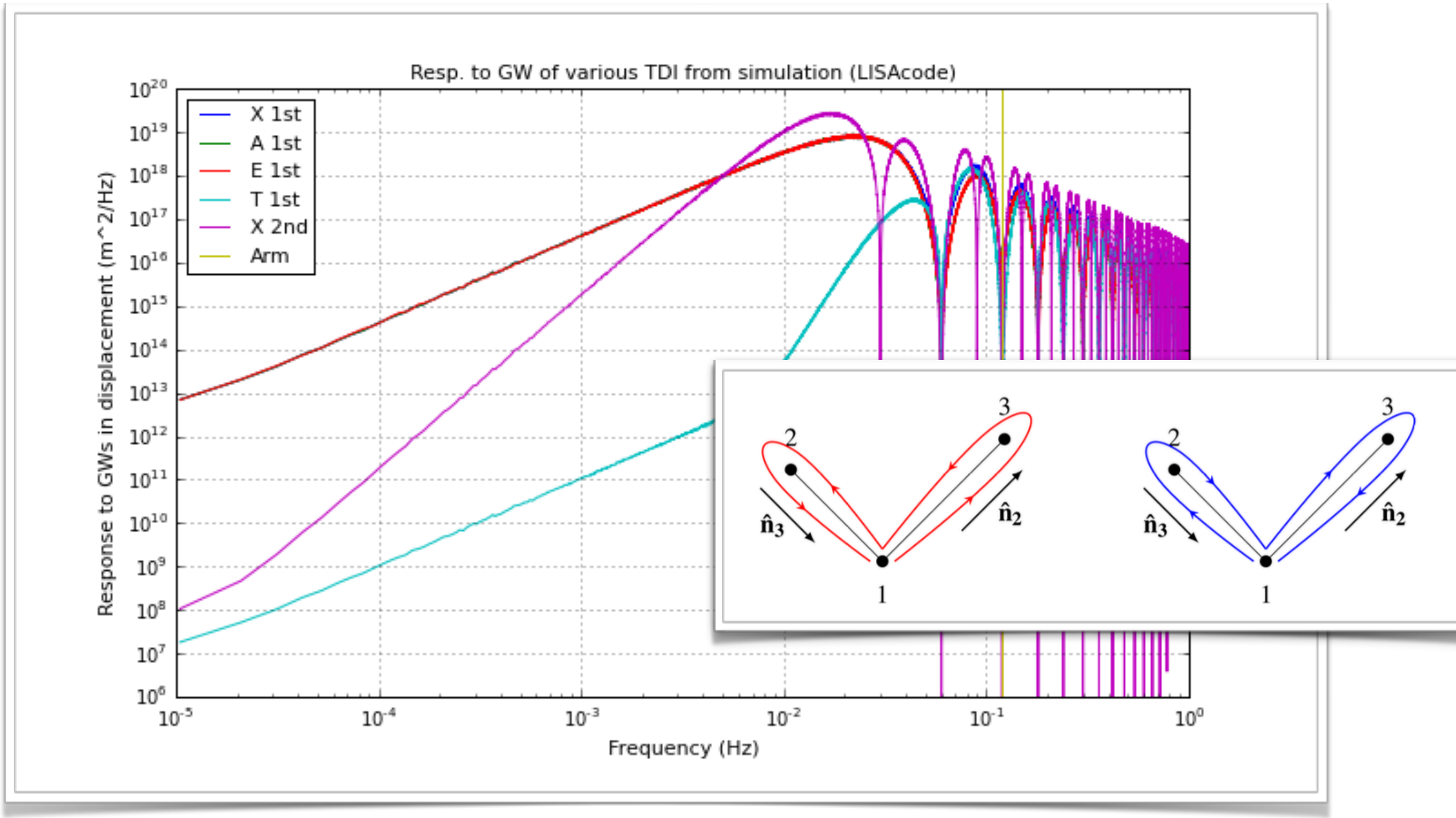
https://lisa-ldc.lal.in2p3.fr/ldc

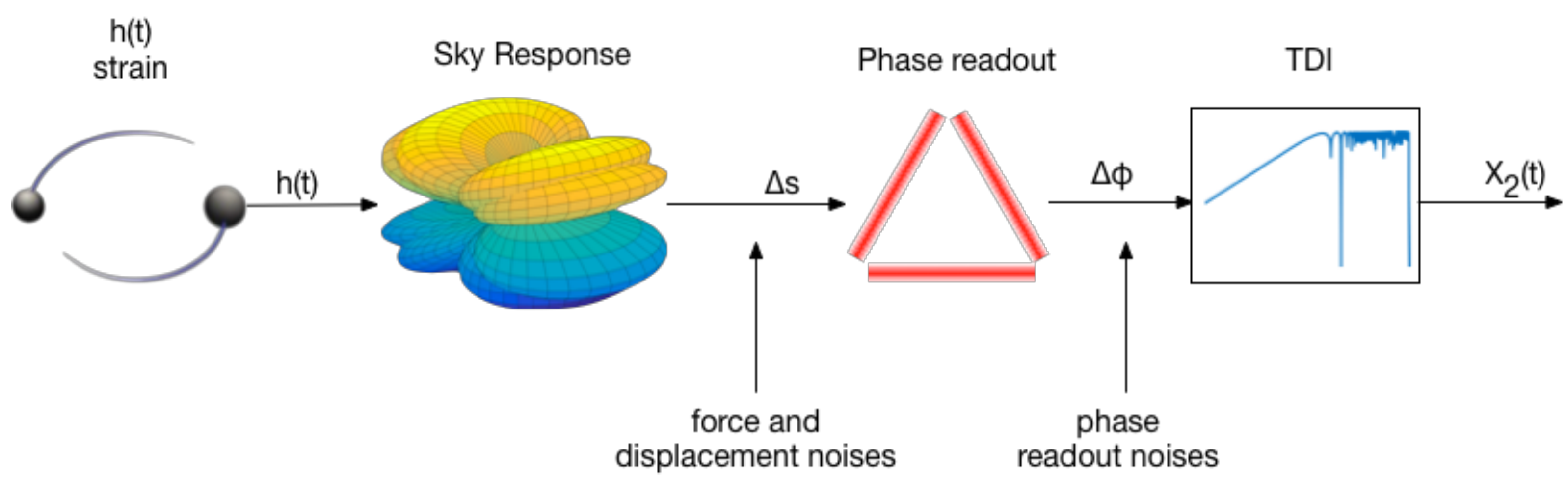
Extra Material



- TDI channels

Laser Interferometer Space Antenna





LISA Data Challenges - The Past

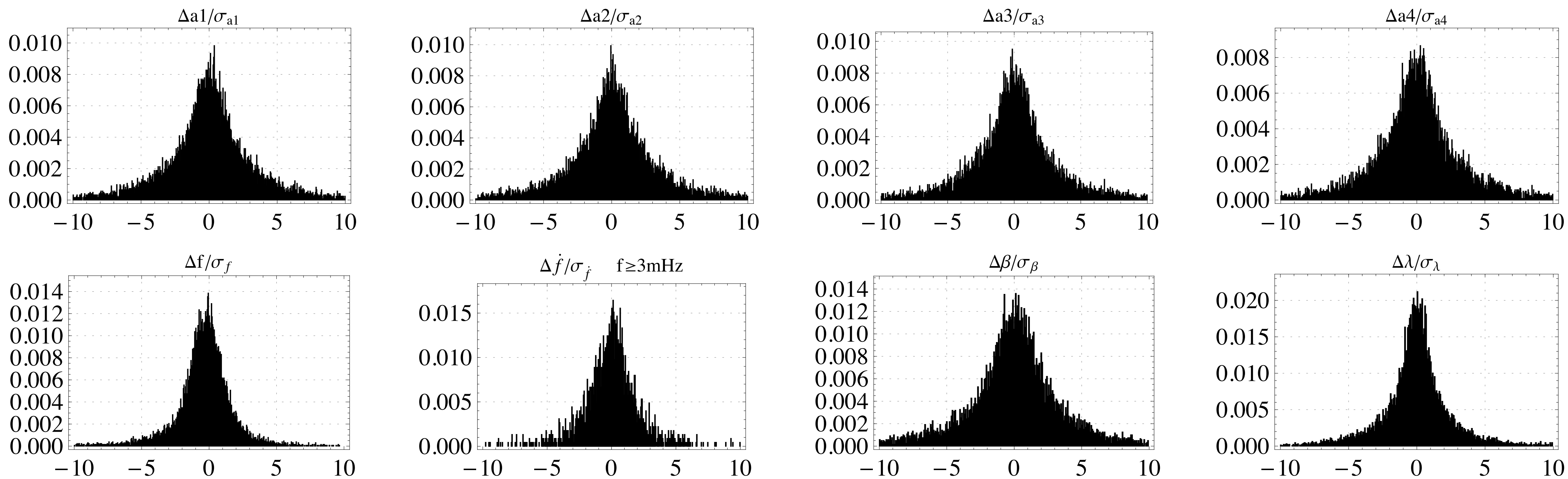
- The results for EMRIs:

Source ($\mathrm{SNR}_{\mathrm{true}}$)	Group	SNR	$\frac{\Delta M}{M}$ $\times 10^{-3}$	$\frac{\Delta \mu}{\mu}$ $\times 10^{-3}$	$\frac{\Delta \nu_0}{\nu_0}$ $\times 10^{-5}$	Δe_0 $\times 10^{-3}$	$\Delta \lvert S \rvert$ $\times 10^{-3}$	$\frac{\Delta \lambda_{\mathrm{SL}}}{\lambda_{\mathrm{SL}}}$ $\times 10^{-3}$	Δspin (deg)	Δsky (deg)	$\frac{\Delta D}{D}$
EMRI-1	MTAPCIOA	21.794	5.05	3.29	1.61	−5.1	−1.4	−19	23	2.0	0.07
(21.673)	MTAPCIOA	21.804	−0.06	−0.01	−0.08	−0.05	0.02	0.54	3.5	1.0	0.13
EMRI-2	MTAPCIOA	32.387	−3.64	−2.61	−3.09	3.8	0.87	12	11	3.7	3×10^{-3}
(32.935)	BabakGair	22.790	33.1	−19.7	10.1	−33	−7.3	250	47	3.5	−0.25
	BabakGair	22.850	32.7	−20.0	9.94	−32	−7.2	250	58	3.5	−0.24
	BabakGair	22.801	33.5	−19.5	10.5	−33	−7.4	240	40	3.5	−0.25
EMRI-3	MTAPCIOA	19.598	1.62	0.38	−0.10	−0.35	−0.94	−3.0	5.0	3.0	−0.04
(19.507)	BabakGair	21.392	1.77	1.01	1.95	−1.2	−0.68	−2.3	116	4.5	0.13
	BabakGair	21.364	2.26	1.88	2.71	−2.0	−0.69	−2.5	65	6.1	0.14
	BabakGair	21.362	1.51	1.01	2.09	−1.3	−0.50	−1.7	7.6	6.2	0.14
	EtfAG	—	54.0	4.88	−7375	26	17	—	—	32	0.83
EMRI-4 (26.650)	MTAPCIOA	−0.441	−8.77	−10.1	−6.03	−3.7	144	950	99	13	−2.3
EMRI-5 (36.173)	MTAPCIOA	17.480	−3.32	5.00	−1.80	0.22	55	62	43	1.8	−1.3

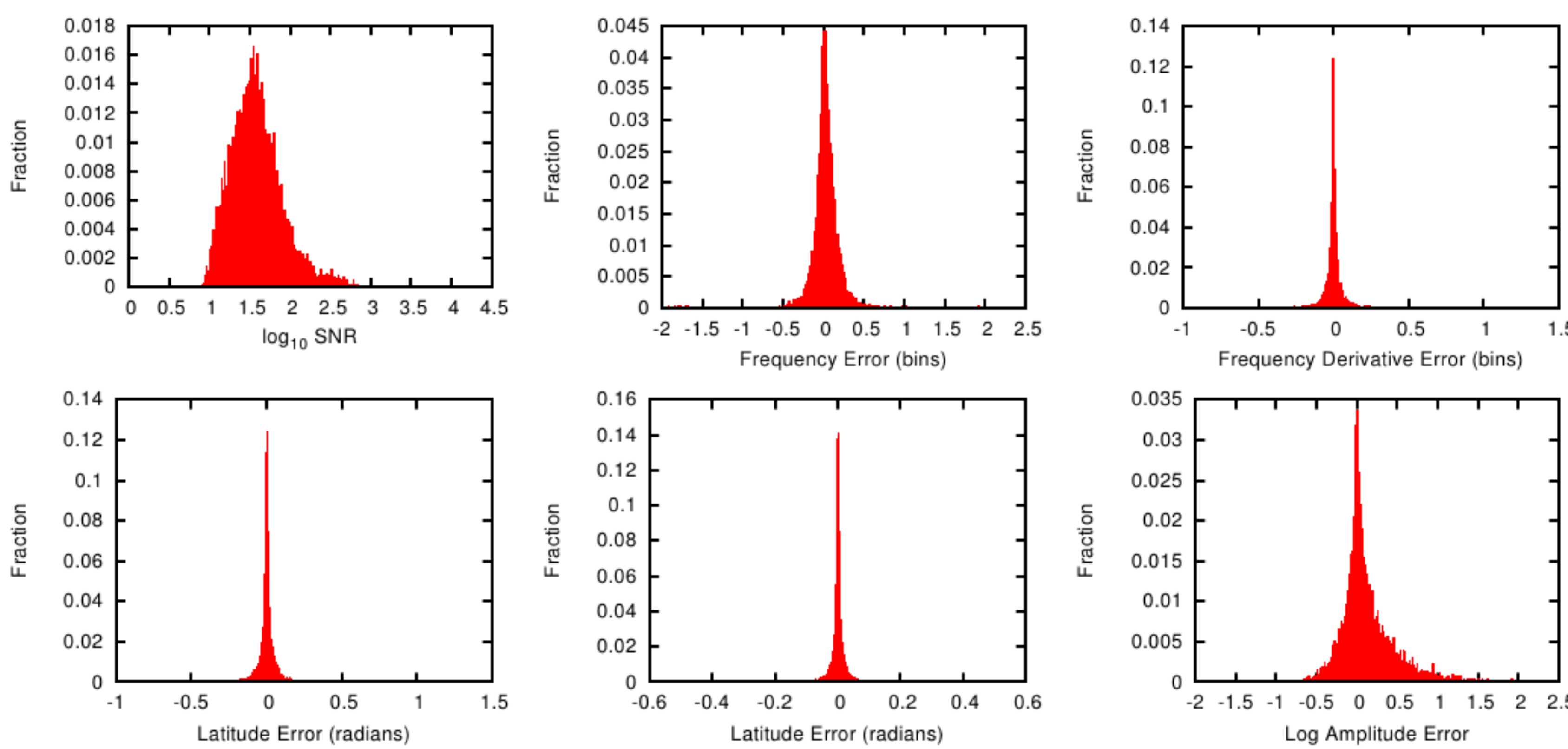
Babak et al, arXiv:0912.0548

LISA Data Challenges - The Past

- The results for Galactic Binaries:



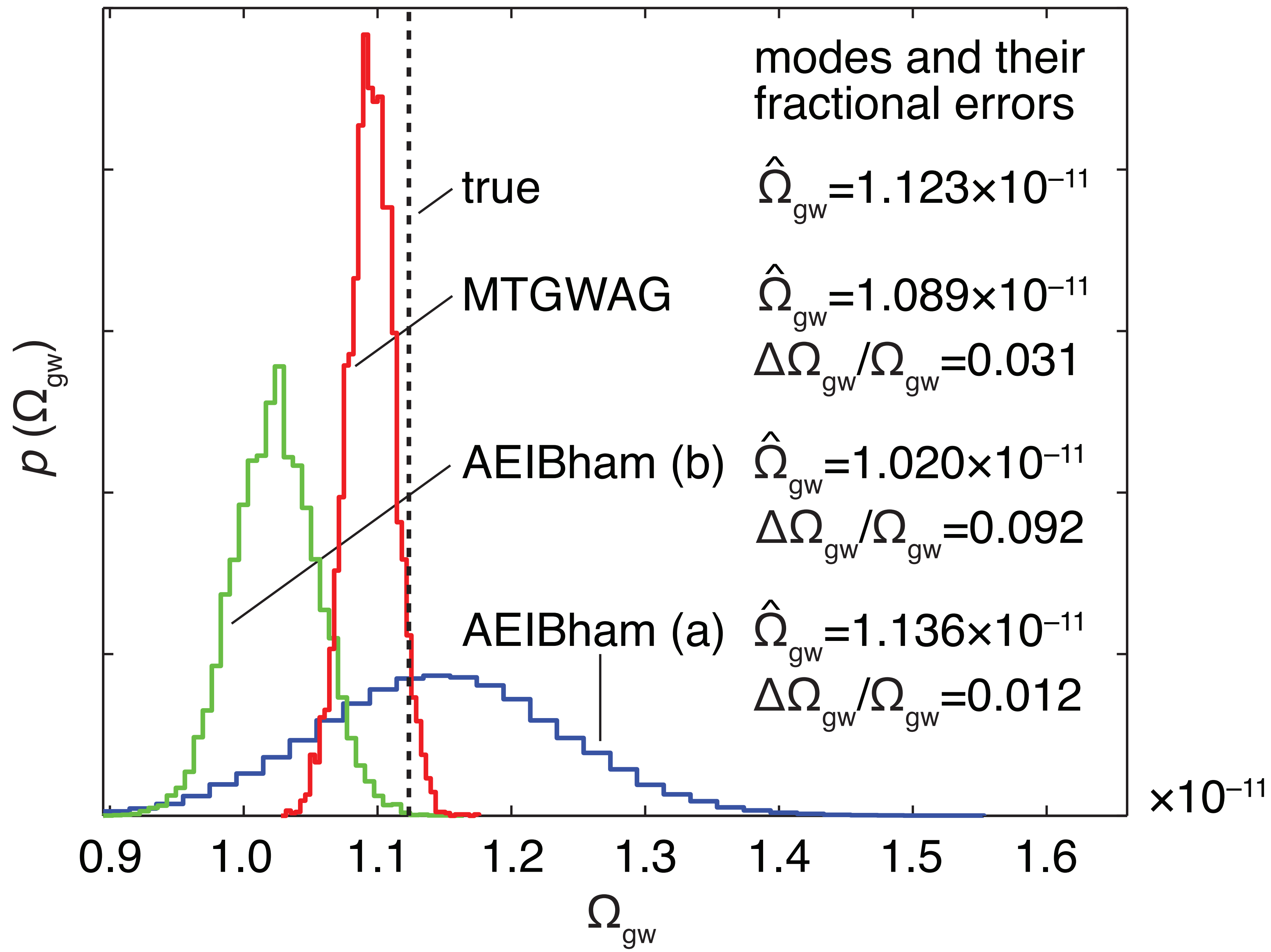
Blaut, PRD 81 063008 (2010)



Littenberg, PRD 84 063009 (2011)

LISA Data Challenges - The Past

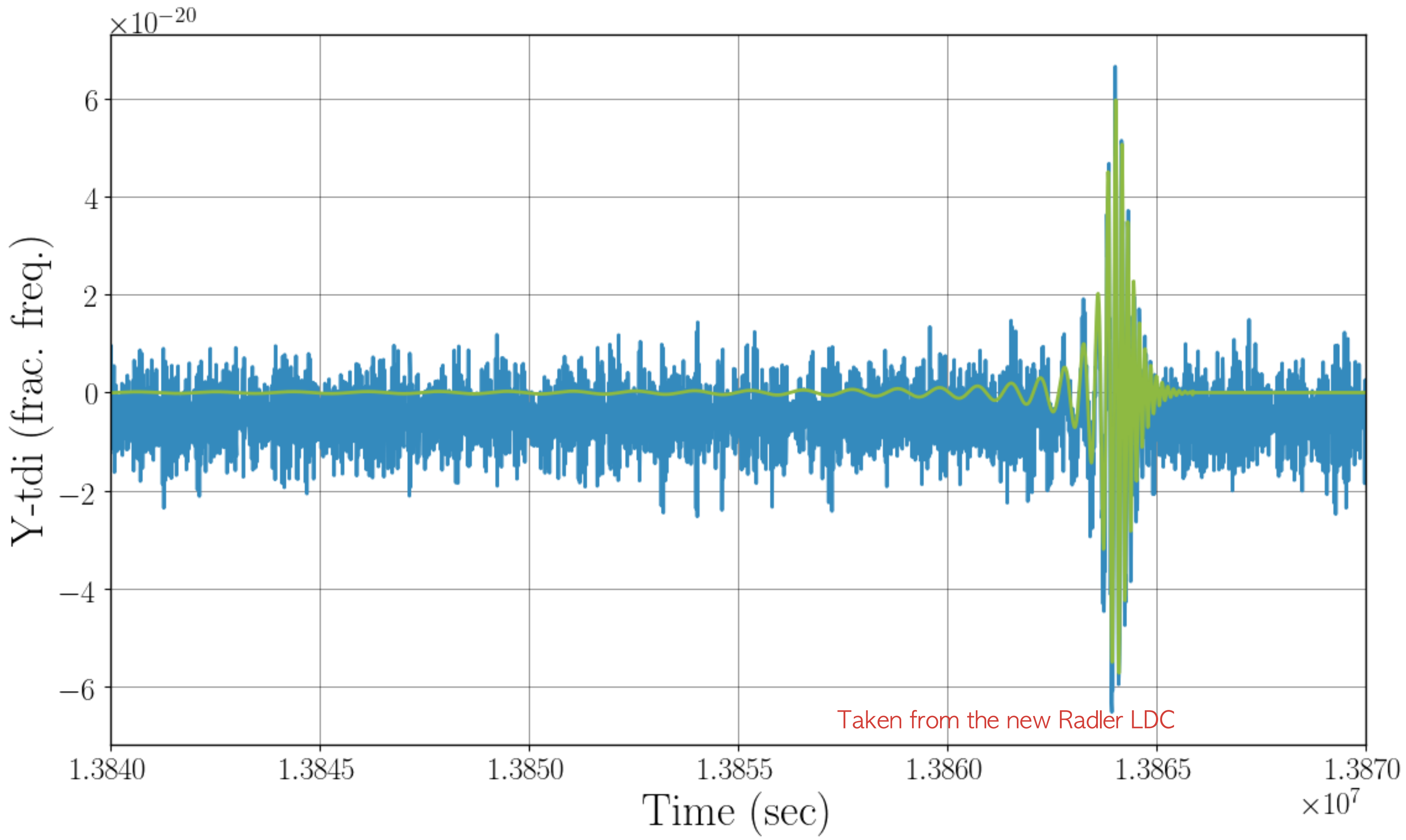
- The results for SGWB:



Babak et al, arXiv:0912.0548

Sources: MBHB

- Expect events with high SNR
- Signal duration varies from weeks to ~year.
- Typically ~15 dimensions parameter space.



http://gmunu.mit.edu/

Data Analysis

- Working in a Bayesian framework we have to explore the posterior probability of the parameters.
- The posterior is defined as:

$$p(\vec{\theta}|d) = \frac{p(d|\vec{\theta}, \mathcal{M})p(\vec{\theta})}{p(d|\mathcal{M})}.$$

- where $p(\vec{\theta})$ encapsulates the prior information we might have.
  - (or not have, so we use non-informative priors)
- The evidence is calculated as

$$p(d|\mathcal{M}) = \int p(\vec{\theta}, d|\mathcal{M})p(\vec{\theta})\mathrm{d}\vec{\theta}.$$

Data Analysis

- We need to measure a stochastic signal,
- where it essentially is buried in the noise of the instrument.
- In general we can assume that we measure

$$d(t) = s(t, \vec{\theta}) + n(t)$$

Then we can approximate

$$p\left(n\right) = C \times \exp\left(-\frac{1}{2}\left(n|n\right)\right)$$

- where

$$(a|b) = 2\int_{0}^{\infty} \mathrm{d}f \left[\tilde{a}^*(f)\tilde{b}(f) + \tilde{a}(f)\tilde{b}^*(f)\right] / \tilde{S}_n(f)$$