Introduction to LISA data analysis

What LISA will observe
Data analysis context
Classical methods and tools

Maude Le Jeune

November 21st, 2022, Toulouse
LISA data analysis: from classical methods to machine learning
Introductory lectures

AstroParticle and Cosmology
IN2P3

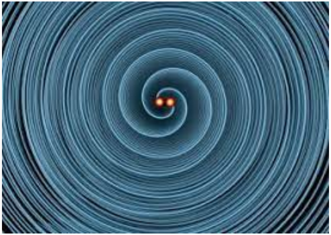
Contents

LISA in the GW spectrum

Contents

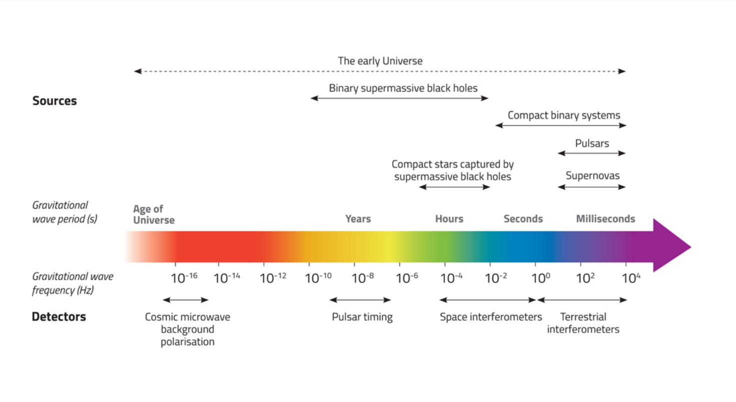
What GW are

- GW are elastic deformation of the space-time metric, they propagate at speed of light
- GW have 2 polarizations (ie quadrupole wave) generated by assymetric matter motion
- GW have been predicted by GR, and first observed by LIGO in 2015
- astrophysical GW source: object needs to have a non spherical acceleration and to be massive/dense: binary systems or asymmetric explosions
- Observational effect: variation of the light-distance between 2 masses at rest: $\delta_l/L = h/2$ with h: the gw amplitude. In the LISA context, we seek for $h \sim pm/Mkm \sim 10^{-21}$



(Image credit: Shutterstock)

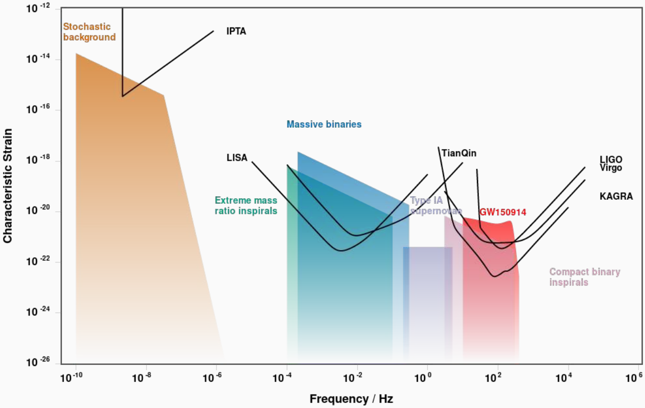
Sources along the GW spectrum



(Image credit: NASA)

- We look for massive / compact binaries
- The wave or orbital period is given on x-axis, by descending order (or increasing frequency)
- Evolution time on a given observable frequency band can differ from millions of years (white dwarf binary) to milliseconds (stellar black hole).
- On the right hand side: lighter (thus nearby) objects
- On the left hand side: heavier, far objects (extra-galactic).

GW observatories: from 2022 to 2035



(http://gwplotter.com/)

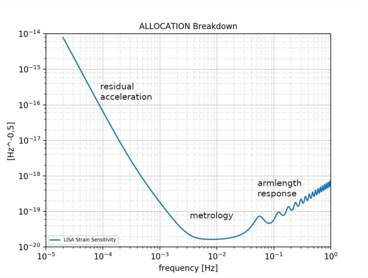
iPTA (international PTA: Europe, North America, Australia, India)

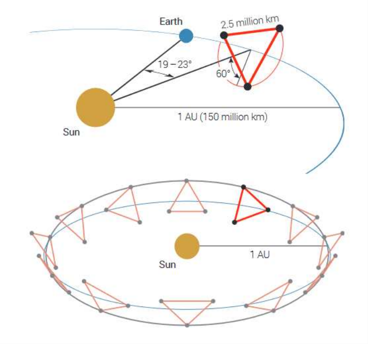
- target galactic millisecond pulsars, L = distance to earth.
- Stochastic background of supermassive black hole binaries $10^9 M_\odot$, up to z=1

LVK (LIGO/Virgo/Kagra)

- $\sim$ 90 events since 2015 (mainly stellar black holes, + some neutron stars mergers).
- O4 in March 2023, almost x 2 in sensitivity, 1 yr of observation. O5 in 2027+ (factor 2 in sensitivity again)

LISA sensitivity and key numbers





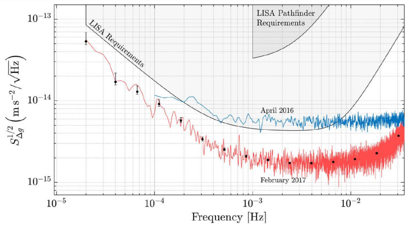
- At low frequency: imperfection of free fall = acceleration noise: require a control the thermal/mag/grav environment. $\sim 3fms^{-2}/\sqrt{Hz}$ around 1-2mHz.
- At high frequency: precision of test mass separation measurement: dominated by photon shot noise, trade-off between telescope diameter (30cm) and arm length (2.5 Mkm) and laser power (2 W) $\sim 12pm/\sqrt{Hz}$
- Above 30mHz: GW period becomes shorter than light travel time, so the signal is partially cancelled out
- Not shown here: the laser frequency noise is mitigated by active stabilization of the laser freq (through a 3rd reference interferometer) and TDI.
- Heliocentric trailing orbits, 20 deg behind the Earth ie 50 millions km, to avoid the thermal/mag/grav perturbations from the Earth.
- Ground communication with only one spacecraft, 334Mb per day (in 8hrs) using 1 to 3 stations of the ESA network, constrained by distance from earth and antenna size.
- Duration: 1yr comissioning + 4 yrs science up to 5 yrs extension

The LISA project

- A space project proposed by the LISA consortium decades ago, selected and led by ESA since 2018, NASA as junior partner, with industrial contractor to be selected in the coming year.

- All technologies must be TRL6 to pass adoption [free fall has already been demonstrated by LISA Pathfinder, with great succes]. After adoption: 10 years to build the instrument and get ready to do a live data analysis.

- Key documents
  - LISA L3 Mission Proposal
  - The science requirement document
  - The redbook (to be released before adoption)
  - White papers, living reviews: references given in the following slides

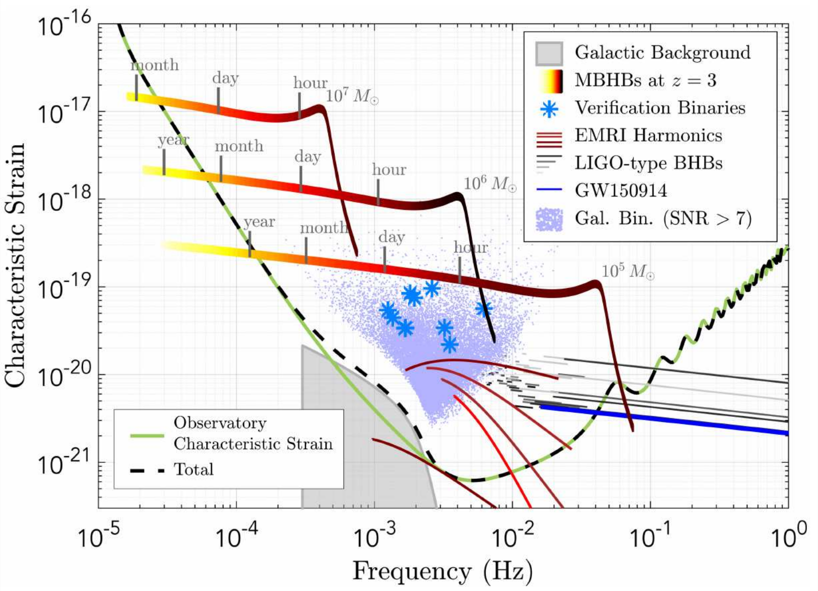


Armano et al., 2018

LISA sources and science case

Contents

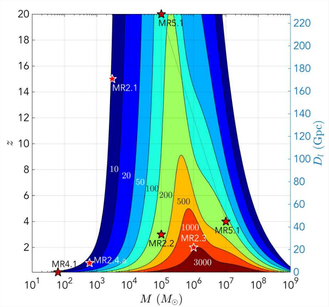
Global picture



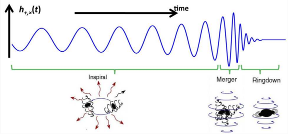
Astro white paper: Amaro-Seoane et al., 2022

Massive Black Hole Binaries (MBHB)

- MBHB are the strongest signal that LISA can see, up to SNR >1000 for low redshift sources.
- GW weakly interact with matter, they travel undisturbed and allow to explore redshift up to z$\sim$20. LISA should see MBHB signal from cosmic to recent time, thus giving insight on the **co-evolution** of galaxies and MBH across cosmic time: formation, growth, ...



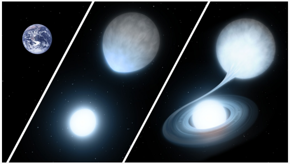
- Expected mass range in $\left[10^4 - 10^7\right]$ solar mass (compare to MBH in Milky way, Sagittarius A, 4×10^6), MBHB are transient signals, LISA will see: late inspiral, merger and ringdown phase.
- Event rate around few - a few hundreds per year.
- GW signal gives access to spin and masses to % accuracy. Spin is difficult to measure with EM spectra. Masses and spins constrain formation scenario: accretion (higher spin) versus merger (lower spin).
- GW signal is the only probe for inactive/quiet objects, so far we only have access to EM signal of active galactic nuclei.



Galactic Binaries (GB)

- In our Galaxy: most of the stars have a companion, half of which is sufficiently close to interact. This leads to millions of sources that LISA could see. Among them: thousands resolvable, rest is a background (a limiting one).
- some ($\sim$ 30) verification systems, already known by EM emission (Gaia, ZTF)

- White dwarfs (most of them), thousands of detached systems or interacting ones (refered as AMCVn where there is mass transfer) but also hundreds of binary with neutron stars (aka UCXB, with X ray emission), BH (10 or 100 of them). In one word: systems at the end state of stellar evolution.
- Monochromatic, persitent objects, period of few minutes. Seen by EM observation, but with extinction (thus difficult to get in the galactic halo). One could use EM (Gaia: few hundreds, LSST: thousand, ZTF, JWST, ...) to refine localization, mass, inclination, orbital decay ($\dot{f}$) : due to GW radiation vs mass transfer.



credit: Caltech

GB science case

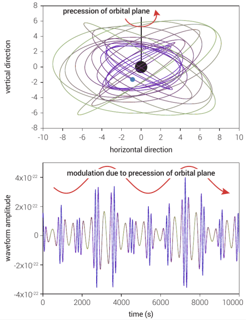
- Thousands of sources with distance and sky position will help refinig our galaxy morphology model.
- Stellar population models (formation+evolution) from coalescence rate.

Stellar Origin Black Hole binaries (SOBH)

From inferred rate of merger discovered by LVK: up to $\sim$ 25 SOBH could be seen by LISA, some of which merging in the LVK band (months to years later). By giving access to the inspiral phase of the signal, LISA should be able to measure eccentricity (binary becomes circular closer to merger).

Extrem mass ratio inspiral (EMRI)

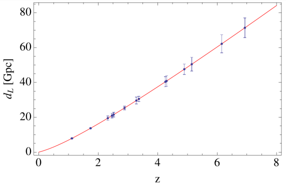
- EMRI are made of a compact star (neutron star or BH) of $\sim$ 10-60 $M_\odot$ captured by BH (of $10^5 - 10^6\ M_\odot$), spiralling in the few Schwarzschild radii region from event horizon (during months to years), then plunging into it.
- Number of cycles is roughly inversely proportional to mass ratio.
- Because of relativistic effects, the waveform is very complex, but we expect to get a lot of cycles (10^4-10^5) thus a potential precise measurement of intrinsec parameters like spin.



  - Spin of the MBH tells about formation history: gas accretion (higher spin) versus mergers with other MBH (lower spin). Eccentricity and orbital inclination gives information on formation of the binary.
  - Also from the long time spent close to the BH horizon, one can measure the multipolar structure of the BH encoded in the GW signal, allowing for tests of GR (no hair theorem).
  - Event rate is uncertain due to the lack of EM observations so far (1-1000 per year). There are many plausible formation channels, such that even no detection would put contrains on them.

EM counterparts and standard sirens

- Ultra compact binaries are guaranted source with EM counterpart in optical and X-ray.
- Regarding BH, if merger happen not in vaccum but in region with surounding gas, EM signal expected (mostly X-ray from gas accretion), at pre-merger, merger, and post-merger (disc rebrightening, corona, jet).
- Radio signal is also expected from AGN relativistic jet.
- Multi-messenger studies involve: LSST ($10deg^2$ FoV), SKA, Athena, ...
- Today: some effort ongoing to better understand EM signatures of MBHB (with numerical simulation), as well as instrumental effort to find observable candidates (highlighting the binary nature of objects in optical / spectro catalogs).
- Alerts with LISA: real time measurement of EM counterpart will bring unvaluable information on the BH environment.
- Offline analysis are powerfull too (evolution, co-evolution studies).



(Tamanini et al., 2016)

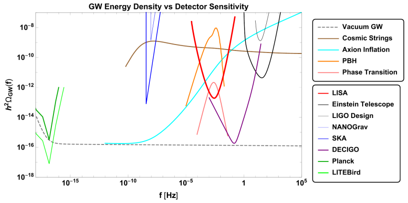
Cosmology with standard sirens

- absolute luminosity $D_L(z)$ encoded in the signal, calibration given by GR
- with an independant redshift measurement, provided by EM counterpart, one can get points on the distance vs redshift curve, and then constrain H_0 at the % level.
- without EM counterpart: one can match sky localization + distance with existing galaxy catalogs, giving several host candidates for each GW event: dark sirens

Stochastic GW Background (SGWB)

of cosmological origin

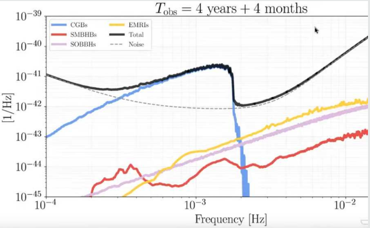
- Cosmological stochastic background of gravitational waves can be produced by several processes in the early universe (like first order phase transition, cosmic string, inflation). Especially first order phase transition is right on the LISA band.
- Measuring such a cosmological background (its shape and amplitude) would allow to constrain theoritical scenarios. Detection of a deviation from isotropy, Gaussianity and circular polarization would also be a valuable interest.



Auclair et al., 2022

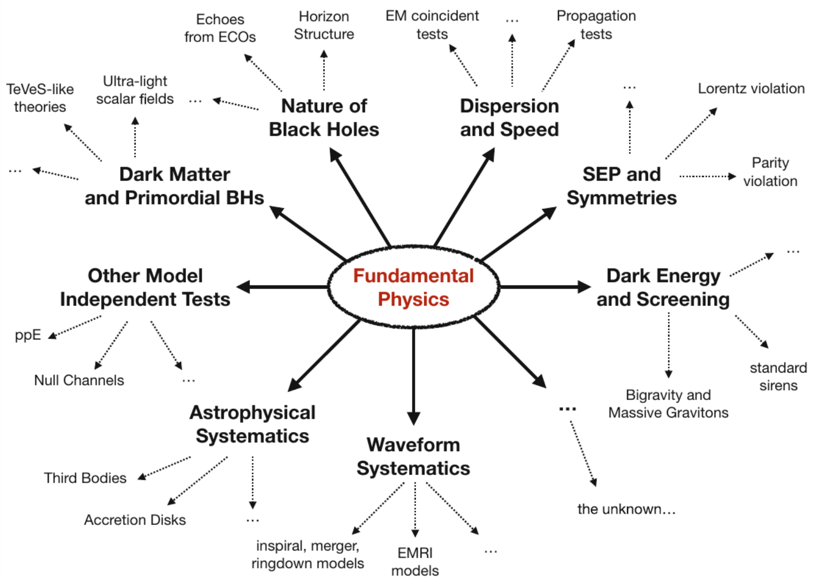
of astrophysical origin

- from any of the source type mentioned so far (SMBH, SOBH, EMRI, GB): carry information on the underlying population
- from unresolved SOBH: spectral index 2/3, amplitude depends on merger rate, and parameter distributions (masses, spins, redshifts). Expected SNR around 15.
- from EMRI: depending on population model, could be negligible to limiting in the $10^{-3} - 10^{-2}$ Hz band.



(Image credit: Karnesis)

Fundamental physics



See living review in relativity Arun et al., 2022

LISA data processing

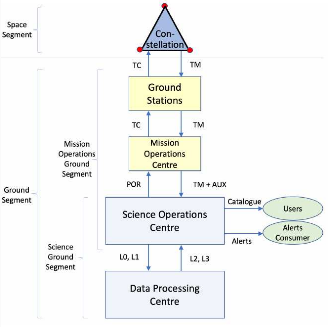
Contents

LISA data flow [1/2]

LISA objectives from Lammers et al., 2019

- All-sky survey of GW sources and creation of a corresponding catalogue
- Issuing of alerts of upcoming GW events to enable contemporaneous observations in the EM band

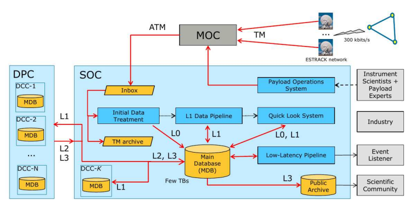
- Ground segment includes: MOC (ESOC/Darmstadt) + SOC (ESAC/Madrid) + DPC (Consortium)
- MOC will receive telemetry data (TM), and send them to SOC.
- SOC will process/complete them to build L0 data, and derive L1 (daily). SOC will deliver LISA end products (catalogs, alerts) to the community.
- DPC will build the catalogs (L1 to L3), produce simulated data, and will be involved in alerts and preprocessing.



LISA data flow [2/2]

- L0 includes telemetry data: all data accumulated by the S/C compressed, packetized, transmitted to SOC via MOC as files.
  - Housekeeping data: sensors monitoring, star tracker, clocks data
  - Science data: phasemeters measurements (long-arm, TM, reference), and all other instruments monitoring (laser, GRS, ...), DFACS data.
- but also: orbits data, auxiliary timing from ground, meteorological data.
- De-packetized, de-compressed, cleaned from corrupted items, converted from row ADC to physical units, clock synchronized, time ordered (L0.5).

- L1: calibrated and noise corrected data
- L2: fully processed data.
- L3: fraction of L2 to be released to the community (catalogue, sources strain time series, residual data, TBD)



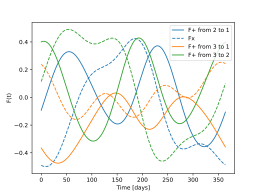
Public data

- Any released item should be reproducible from L1: algorithm, software, models need to be released too. Discussions on the potential release of L0/L0.5 ongoing.
- Data will be given in open source / standard format, software with open source license. A support will be provided by SOC+Consortium.

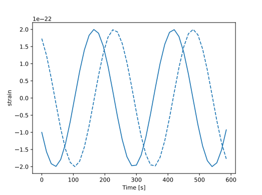
Detector response to a GW source strain

- The GW strain is a dimensionless number characterizing amplitude of spacetime streching caused by GW: $h(t) = \delta_L(t)/L = \delta_t(t)/(2L/c)$
- The detector response is the phase difference in the laser light between 2 arms of the interferometer, or equivalently the laser light travel time difference $\delta_t(t)$
- Because the GW signal is weak, detector response is linear in the wave metric perturbation, $h(t)$ is the convolution of the metric perturbation with the impulse response of the detector.
- Since it's a polarized signal, we derive h_+, h_x (the waveform) in source frame, then detector frame (solar barycentric system for LISA), then apply antenna response $F_{+/x}(t)$.
- It varies with time following orbital motion of the constellation, and is link dependant.
- Sensitivity to a given GW source depends on the orientation of the constellation wrt to the source, which changes with time.

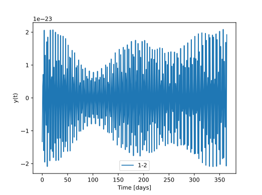
$F(t)$ over 1 year



$h(t)$ over 10 min



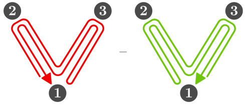
$y(t)$ over 1 year



Time Delay Interferometry

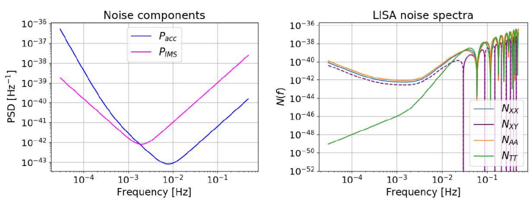
- Laser noise is 8 orders of magnitude above the GW signal, and thus should be cancelled. $y_{rs}(t) = y_{rs}^{GW}(t) + y_{rs}^{laser}(t) + y_{rs}^{acc}(t) + y_{rs}(t)^{oms} + \ldots$

- TDI combines the delayed interferometric measurements along the 6 links to cancel the laser noise (of the 3 s/c), and build 3 TDI variables. 1*st* generation was derived for a unequal but fixed constellation (8 terms), 2*nd* generation works with time varying armlengths (16 terms).



(Image credit: J.-B. Bayle)

- There is an infinity set of combinations which allows to cancel the laser noise. Once chosen, one needs to derive the response of GW and other noise source to it.
- Usual combination is Michelson variables (X,Y,Z), which gives virtual 3 Michelson equal arm interferometers measurements, but correlated noise. The combination widely used in data analysis is the un-correlated AET decomposition Prince et al., 2002 , with T less sensitive to GW signal, can be seen as a null channel.

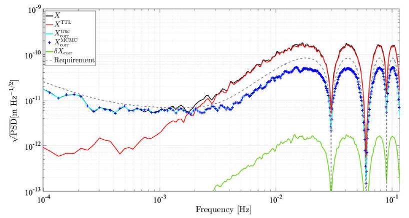


Auclair et al., 2022

Other noise reduction and calibration

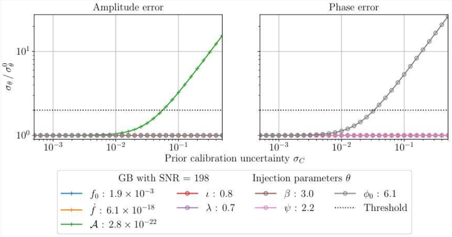
Noise reduction

- Full L0 to L1 preprocessing includes a various other noise reduction in addition (or as part of) TDI
- TTL noise: relative jitter between OB and distant laser beam, change optical path length. TTL correction can be done using on-board DWS measurements (differential wavefront sensing) and a TTL coefficients fitting procedure. Paczkowski et al., 2022



Calibration

- Rely on our knowledge of the wavelength of the laser light and onboard clocks timing, and translates into a calibration of signal phase and amplitude
- In order to keep the calibration error smaller than the random error, especially for distance measurement, we need $\sigma_a < few 10^{-3}$ (and $\sigma_\phi < 10^{-3}$) (Savalle et al., 2022).
- Sources with known luminosity distance (like VGB) all together could be used as cross check, with a precision of few % in amplitude. Phase calibration could be checked with EMRI.



LISA data analysis

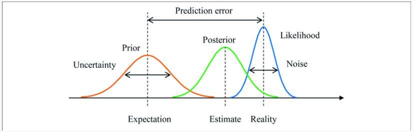
Contents

Concepts and tools

GW data analysis

GW event detection and characterization lie on a hypothesis testing framework (or inference)

- Hypothesis are usually based on a model with some parameters θ
- Question is: does the data d favors one hypothesis or the other ?



Frequentist approach

$p(\hat{d}|\theta) \rightarrow$ the likelihood

- A true value for the parameters exists, the data is a particular noisy realization of the model.
- One use an estimator $\hat{\theta}$ which can be the maximum likelihood
- If model is correct, one can derive confidence intervals or p-value and lower bound uncertainties on the best-fit parameters from the Fisher Information Matrix

Bayesian approach

$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)} \rightarrow$ the posterior

- Given the data, what is the probability for the parameter to take a particular value.
- $p(\theta)$ reflects prior knowledge on the model: ranges and distributions
- The posterior gives the most complete information about the parameters, one can derive confidence intervals, maximum a posteriori, ...

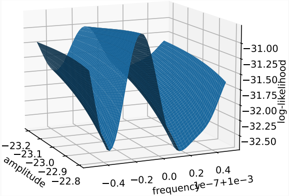
Likelihood in GW search

- Assuming independance of the signal with respect to the instrumental noise: $d(t) = s(t) + n(t)$
- and that the residual $r(t) = d(t) - h(t) = n(t)$ is left with Gaussian noise (by central limit theorem)
- Likelihood is given by the Gaussian distribution: $p(d|h) = \frac{1}{\sqrt{\det(2\pi C_n)}} e^{-\frac{1}{2}\sum_{i,j} r_i \left(C_n^{-1}\right)_{ij} r_j}$
- C_n is the noise correlation matrix, diagonal in Fourier domain for a stationary noise: $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{r^2}{2\sigma^2}}$
- For multiple detectors, the sum goes over both samples and channels (like A, E if working with TDI AET decomposition).
- Matched filter based likelihood can be used to find a good starting point. It is based on SNR maximization, and involves inner product: $(d|h) = SNR^2 = 4\Re \int_0^\infty \frac{\tilde{d}(f)\tilde{h}(f)^*}{S_n(f)} df$
- In practise, we use the log-likelihood: exponential likelihood, normal distribution is log-concave, products become sums, and other pratical reasons.

Parameter fitting

One is mostly interested in getting a small residual, such that a big part of the work if about optimization.
Multiple techniques are broadly used in data analysis:

- Gradient descent methods (w/wo preconditioner): first/second order iterative optimization of the likelihood.
- Stochastic methods (Monte Carlo Markov Chain, Nested Sampling): not strictly speaking designed for it, but good at finding high density region of the posterior while sampling.

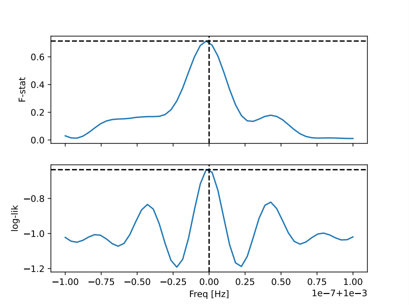


In GW data analysis

- High dimensionality of the model and multi-modality of the likelihood/posterior are the main issues in this optimization problem.
- The 1st category also suffers from the need to keep parameters in their physically allowed ranges, which creates discontinuity of the likelihood surface.
- Most GW data analysis falls into the 2nd category.

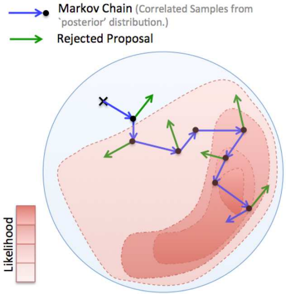
Reducing the dimensionality with F-statistic

- In GW data analysis, F-statistic can be used to reduce the dimensionality of the optimization problem, by splitting them into intrinsec (related to the source itself) vs extrinsec (related to the observer).
- Going back to the expression of the response function of the detector to a GW, one can re-write it as a linear combination of 4 functions: $h(t) = \sum_{k=1}^{4} a^k h^k(t, \xi^\mu)$ where a^k are the extrinsec parameters.
- From this linear combination, one can find a closed form ML estimator $\hat{a}$
- Using this estimator, the log likelihood can be re-written in a form which only depends on intrinsec parameters, which is called the F-statistic.
- F-statistic then can help in solving the ML optimization problem or to find a good starting point.

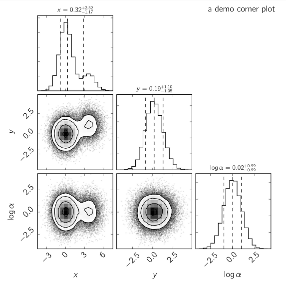


Stochastic methods

Stochastic methods aim at mapping the posterior distribution, by randomly generating a set of samples θ_i (a chain), iteratively, until they faithfully describe the true posterior distribution.



From https://www.turing.ac.uk



From https://pypi.org/project/corner/

Different approaches exist:

- Metropolis Hasting sampling and its extensions (multiple try or delayed rejection, tempering)
- Hamiltonian sampling: replace random walk by state space proposal
- Slice sampling: sample from the area under the graph of the density function
- Splitting the parameters to look iteratively for marginal distributions instead of the full joint distribution (Gibbs sampling)

A special case is the Nested Sampling techniques, which aim at estimating the evidence, and for which the posterior is a by-product.

MCMC [1/2]

How it works

- Key requirements to ensure the convergence to target posterior distribution are:
  - irreducibility: all points should be visitable in a finite number of steps
  - aperiodicity: one should not get trapped in some cycle

  which in turn constrain the decision rule of accepting/rejecting new samples.
- The baseline is to use the Metropolis Hasting acceptance probability: $A(\theta_{n+1}, \theta_n) = min(1, \frac{p(\theta_{n+1})p(d|\theta_{n+1})}{p(\theta_n)p(d|\theta_n)})$ where new points are randomly chosen from $p(\theta)$.
- A higher likelihood increases the probability of being accepted.
- One can increase the probability for new points of being accepted by learning from the chain history: adaptive MCMC. This breaks the Markovian property of the chain, but assyptotically, one assumes to get a Markov chain back once everything's stabilized.

Convergence

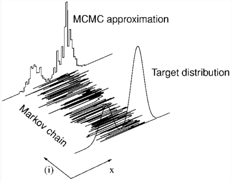
- Far from the high density region of the posterior, it is difficult to propose new samples which are accepted, this increases the auto-correlation length. Usually burn-in part of the chain is discarded.
- On the other hand, one can design a stopping criteria from the effective sample size (ESS).
- Highly correlated (degenerated) parameters are of the main reasons for a slow convergence (requires time to learn the correlation and propose good points), and should be avoided whenever possible.
- Another one is when the chain get stuck in a secondary maximum.

MCMC [2/2]

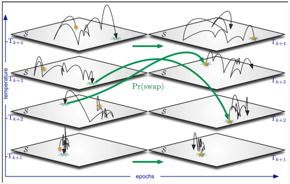
Ensemble chains

A common way to use MCMC is to run chains in parallel:

- N chains gives N more samples of the posterior for a given time. Mixture of chains can improve the proposals.
- Having several chains running in parallel offers the possibility to use parallel tempering with different cooling at the same time $p(d|\theta) \to p^{1/T}(d|\theta)$
- By flattening the likelihood, high temperature chains can explore all the modes of the posterior, while the coolest ones focuses on the main one. Point exchanges between the chains triggered by a higher likelihood, allows newly discovered modes to become the important ones.
- This helps to solve the secondary maxima issue.
- In the case where several chains are run in the same time, the Gelman-Rubin ratio is commonly used to assess the convergence. It is based on the ratio of the within-chain and between chain variance.
- This can be extended beyond the single parameter comparison, using distance (like Kullback-Leibler) of the chain's KDE.



Andrieu et al., 2003

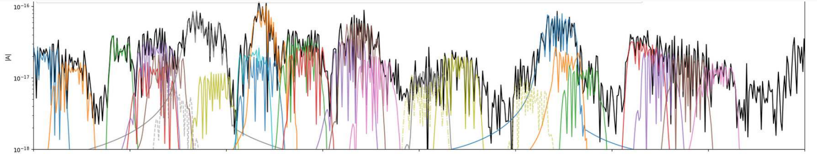


Jacobs et al., 2015

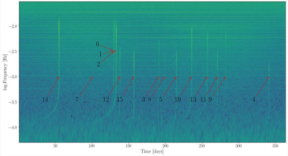
The global fit

LISA sources and the global fit problem [1/2]

- One needs to infer the parameters of GW sources of different kind, with different properties
- GB: thousands of sources to find, 8 parameters each. Model selection is needed here.
- BHB: not too many sources but some tricky parameter degeneracies and correlations.
- GB are quasi monochromatic event → narrow band in the frequency domain: one can split the frequency domain into independent regions where to look for a reduced number of sources at the same time.

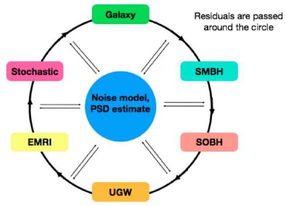


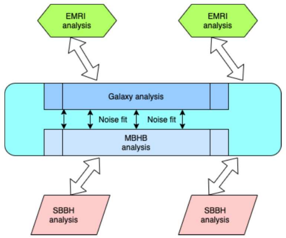
- On the other hand BBH are transient signals, with response varying both with time and frequency.
- Splitting the source identification into 2 independant steps would simplified the problem a lot.
- But, one needs to fit them together in order to properly account for their respectives uncertainties (or removal residuals) impacting one with each other.



LISA sources and the global fit problem [2/2]

- A careful modeling of the (non Gaussian) residual coming from resolved bright sources is also required to fit for stochastic background, such that SGWB fitting must be part of the global fit too. Auclair et al., 2022
- It is assumed that the joint identification can be done in an iterative way (Gibbs sampling), assuming small correlation between the parameters of the different resolvable source types and background.
- There are some ongoing studies to demonstrate the performance of that Gibbs sampling approach:
  - SGWB / resolvabes sources in the LVK context Biscoveanu et al., 2020
  - GB / BHB in the LISA Data Challenge context
- This requires some load balancing between update of each source type, weighting including: number of sources (GB), speed of convergence (multi-modality, correlations of BHB), speed of likelihood evaluation (waveform complexity of EMRI) etc





Bottlenecks and the need for GPU/ML

Contents

L1 to L3

CPU needs of the global fit

- Cost is driven by the time take to compute a likelihood evaluation and number of likelihood evaluations needed to converge.
- Time to converge is governed by multi-modality of the likelihood, presence of correlated parameters, number of parameters jointly explored.
- Likelihood evaluation is dominated by the waveform computation, for which a broad range of approximations are explored, together with their impact on source identification (bias, variance).
- Varying the likelihood approximation along the MCMC run can be used to mitigate this rough product (F-statistic, heterodyning, machine learning, etc)
- GPU can be used on top of that: parallelization over time or frequency series, or over multiple points in MCMC machinery.
- All those considerations enter the MFR cost study Babak, 2021 with actual numbers supported by LDC R&D.

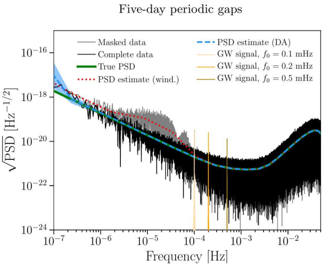
Source type	CPU-hours	Scratch volume	Informative volume
Galaxy	(180-250)K	(260-2000)GB	120GB
MBHBs	(1.2-300)K	(5-50)TB	(0.6-6)GB
EMRIs	(4-6)M	(16-24)TB	(12-20)GB
SBBHs	14M	100TB	200MB
Noise	(0.5-3)K	(50-260)GB	(1-5)GB

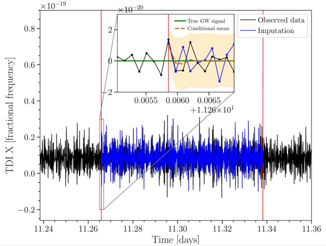
- 1 year, analysis of all resolvable events: $\rightarrow$ **20Mcpu.h to 30Mcpu.h** (French support to the project today is $\sim$ 5Mcpu.h)
- 3 iterations per year, 2 pipelines gives an additional factor 6
- R&D requires to multiply this quantity by a big prefactor...

Glitches, gaps, non-stationarities

Gaps

- Short ($<$ hours) gaps are expected from antenna repointing (9 days) and laser relocking (several weeks), plus unexpected long gaps ($\sim$ days) like in LPF are often considered. Total duty cycle is 80-90%.
- Gaps in time series lead to spectral leakage when switching to Fourier domain, and off-diagonal terms in the noise matrix. Apodization in time domain can reduce the spectral leakage, but inpainting approach is better to solve the 2 issues at the same time.
- Additional parameters can be use to draw some signal in the missing part, then iterate on the GW signal parameters in a blocked Gibbs sampling mode.
- Number of those additional parameters depends on the number of the neighboring samples are used to set condition on the inpainted samples: depends on gap size, then trade-off between accuracy and computing time.
- Works well for GB, short glitches, at the price of spending more time in gap filling than in GW source sampling Baghi et al., 2019 .
- Regarding MBHB, a more challenging task due to their transient nature, but should be OK if protected periods are in place Dey et al., 2021.





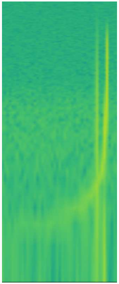
Impact of glitches

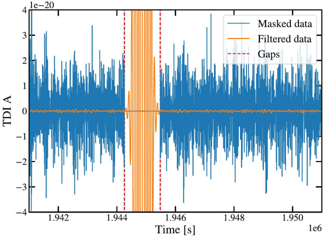
Two kind of issues:

1. High amplitude glitch on top of a transient event
2. High number of glitches preventing from long duration analysis

Both are adressed by the ongoing *Spritz* data challenge, using glitch models from LPF.
Removal in two steps

- Detection: matched filter assuming a known shape and response, or simple low pass filtering.
- Mitigation: Once detected, one can think of different approaches including:
  - Masking + inpainting: model independent approach, but you may loose an important part of the data.
  - Subtraction (Davis et al., 2022 in the LVK context): need to fully understand the uncertainties associated to it and propagate them to GW signal study.
  - Joint fit of glitch and GW signal parameters (Hourihane et al., 2022) provides a proper error propagation, again at the price of an increase of the number of parameters to fit jointly.





(Q. Baghi)

Alerts

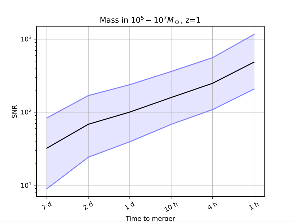
Alerts: key numbers

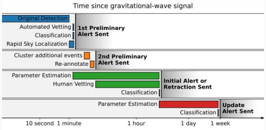
Two timescales envisaged

- Prompt transient events: objective is to have simultaneous EM observation, would require an alert 1 day prior to merger, with refinement on time and localization up to some hours before merger.
- Short term astrophysical phenomena: detection of an inspiral about 2 weeks before merger, would give time to have a protected period after 2 days of 14 days long. During the protected period, science data should be available no later than 2 hours after measurement (using 3 stations of ESA network for a 24/24 downlink)

Low latency pipeline should be able to

- trigger initial alert within hours after L0 data availability.
- refine source parameters before/after merger (especially localization) in a continuous way, with ouptut within hours too.
- LIGO have capabilities of issuing alert within 1 minute with template banks.
- Time constrain is less stringent in LISA, but we are source dominated: less waveform approximations (sky multi modality of MBHB), fast removal of strong contaminant/artifacts, ...
- A time constrained data challenge might help in developing the dedicated toolbox needed for low latency analysis.





https://emfollow.docs.ligo.org

Tools

Contents

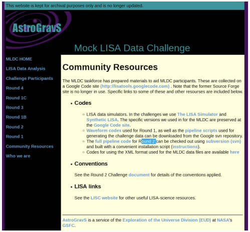
The LDC project

Mock LISA Data Challenge (2006-2010)

4 data challenges have been run to foster the development of data analysis techniques for LISA, with dozens of teams participating.

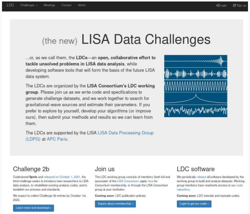
- MLDC1: dozens of GB, 1 MBHB
- MLDC2: millions of GB, dozens of MBHB and EMRIs
- MLDC3: + burst, stochastic background,
- MLDC4: full enchilada

Two simulators were used, waveforms generator for each source type, common convention to distribute the data and collect the results. See dedicated chapter of LDC living review.



LISA Data Challenge (2017+)

- More project oriented, like supporting specific question along LISA design. Less competition, more coordination.
- Some support from the project, to develop software with long term (or broader) support (DPC activity), reproducible pipeline, etc.
- Maintenance of a set of tutorials, to get new people up to speed.



LDC status

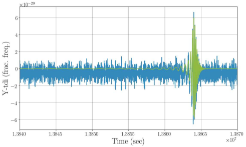
Ongoing challenges

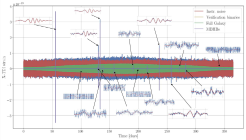
- LDC1/Radler: Verification GB, 1 MBHB, full galaxy, stochastic background, EMRI, SOBH
- LDC2a/Sangria: Full galaxy + dozen of MBHB and VGB Le JeuneandBabak, 2022
- LDC2b/Spritz: Isolated signal with glitches, gaps, non stationary noise, 2nd generation TDI
- LDC1b/Yorsh: Update on EMRI (new waveform) and SOBH.

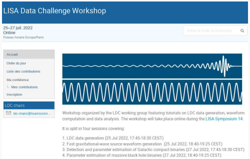
Resources

From the LDC portal:

- data: observations and true signal TDI X,Y,Z
- software
- documentation: conventions, source parameterization, instrumental design
- tutorial notebooks: see gitlab or LDC workshop indico
- + gitlab wiki for ongoing work like living review, etc

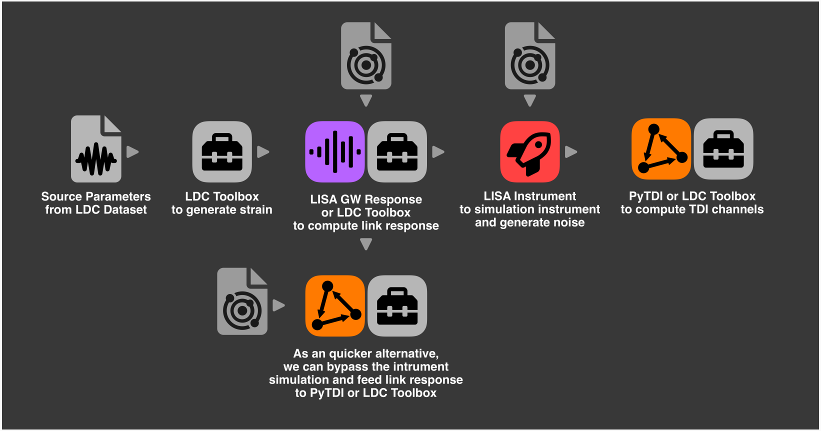






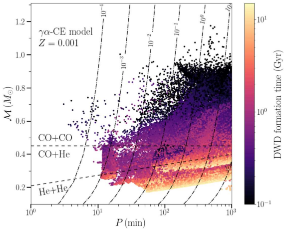
Simulation

Overview



(Image credit: J.-B. Bayle) Simulation is a very active project in LISA, with co-evoluting tools.

Sky simulation



Korol, V. et al., 2020

Relies on:

- Population models, given as one or several realizations (catalogs)
- Randomization of some parameters
- Selection: number of sources, min SNR, etc

LDC catalogs

- An online reference of all known verification systems
- GB interacting from Nelemans et al., 2004, detached from Korol et al., 2020
- MBHB popIII, Q3d, Q3d no delay from Klein et al., 2016

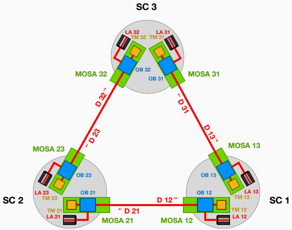
Tools

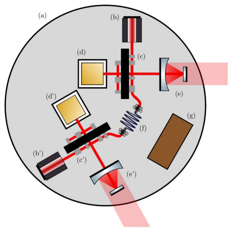
- LDC: Catalog processing
- lisaorbits Bayle, Hees et al., 2022: Orbit file generation (analytic, ESA like)
- LDC: Time domain computation of waveforms, projected strains and TDI (GB, BHB, EMRIs)
- gw-response Bayle, Baghi et al., 2022: Time domain computation of waveforms and projected strains (GB, stochastic background)

Noise simulation

Relies on a model of the LISA constellation:

- generate random noise for each noise source
- combine measurements like LISA will do
- post-processing of the dominant noises: TDI, but not only





Tools

- lisainstrument and lisanode Bayle, Hartwig et al., 2022
- lisaglitch Bayle, Castelli et al., 2022
- pytdi Staab et al., 2022

Additional resources

- Simulation Model for the LISA Instrument
- Tutorials from the simulation workshop

Data analysis

Fast waveforms

Existing waveforms

Fast means: TDI in frequency domain, size of few hundred samples in 1ms.

- LDC provides: GB, MBHB, EMRI FastAK fast waveforms
- lisabeta (Marsat et al) provides: PhenomD, PhenomHM for MBHB, SOBH (existing interface to LDC)
- GBGPU M. L. Katz, 2022 , with GPU support
- BBHx M. Katz, 2021 provides: PhenomD, PhenomHM, with GPU support
- TDI1.5 and TDI2

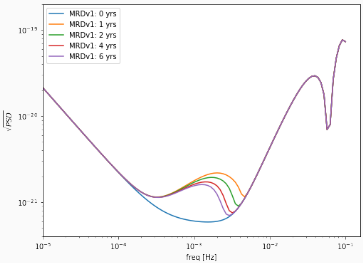
No TDI

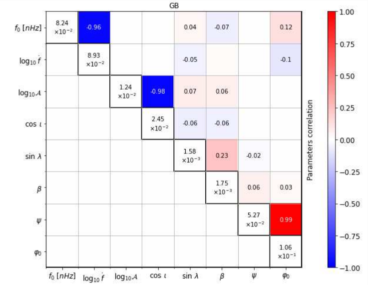
Some waveforms available, but without TDI response

- FEW provides several EMRI waveform with GPU support M. L. Katz et al., 2021
- TBC

DA tools: SNR, FIM, ...

- LDC provides SNR computation facilities
- multi-fisher SAVALLE et al., 2022 provides a common framework to perform FIM computation for all kind of sources, based on LDC
- LDC provides FFT, iFFT wrapper on top of a time/freq series container (xarray), and some windowing facilities
- LDC provides some analytic noise models
- lisabeta provides noise models and FIM facilities for BHB.

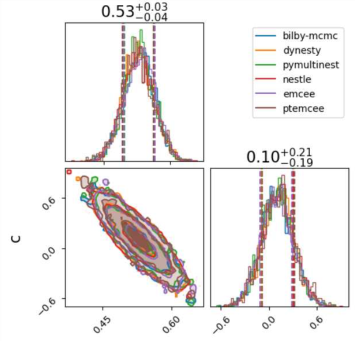




MCMC and NS sampler

Generic and publicly available MCMC samplers

- emcee, ptemcee: commonly used ensemble MCMC sampler
- dynesty, cpnest, nessai: commonly used nested sampler
- bilby: Bayesian framework with interface to a broad range of samplers (dynesty, pymultinest, emcee, ptemcee, pypolychord, etc). See also samplers-samplers-everywhere notebook demo.



It's sometimes useful to spend time in customizing or re-writing those tools to add domain specific features. It's also a good way to get a deeper understanding of how they works.

Summary

- LISA will observe GW sources of different kind, in this new mHz window. The LISA science case is very broad and unique.
- With a laser noise lying 8 orders of magnitude above everything else, the LISA data analysis will require a deep understanding of the instrument systematics, and TDI.
- With very high SNR event of MBHB merger on one hand, and faint potential cosmological background signal on the other, data analysis techniques should be developed and exercised in very different contexts.
- LISA is a source dominated mission. Overlapping sources and the absence of an independent measurement of the noise is a major change wrt ground based techniques used so far. More sophisticated approaches for the data analysis are needed.
- Data challenges have proven to be very efficient in fostering those developments. They now benefit from a very advanced state of the project simulation capacity, and existing tools.
- Adding more and more complexity to the data will reinforce the need for speeding-up existing approaches. GPU and ML are more and more used to that aim, with very promising results.
- The hope is that those new approaches and tools will remain accessible to new comers, both in terms of availability and readibility.

The End

Have a good workshop !

References

Amaro-Seoane, P., Andrews, J., Sedda, M. A., Askar, A., Balasov, R., Bartos, I., Bavera, S. S., Bellovary, J., Berry, C. P. L., Berti, E., Bianchi, S., Blecha, L., Blondin, S., Bogdanović, T., Boissier, S., Bonetti, M., Bonoli, S., Bortolas, E., Breivik, K., . . . Vigna-Gómez, A. (2022). Astrophysics with the laser interferometer space antenna. https://doi.org/10.48550/ARXIV.2203.06016

Amaro-Seoane, P., Audley, H., Babak, S., Baker, J., Barausse, E., Bender, P., Berti, E., Binetruy, P., Born, M., Bortoluzzi, D., Camp, J., Caprini, C., Cardoso, V., Colpi, M., Conklin, J., Cornish, N., Cutler, C., Danzmann, K., Dolesi, R., . . . Zweifel, P. (2017). Laser interferometer space antenna. https://doi.org/10.48550/ARXIV.1702.00786

Andrieu, C., de Freitas, N., Doucet, A. & Jordan, M. I. (2003). An introduction to MCMC for machine learning. *Machine Learning*, *50*(1–2), 5–43.

Armano, M., Audley, H., Auger, G., Baird, J. T., Bassan, M., Binetruy, P., Born, M., Bortoluzzi, D., Brandt, N., Caleno, M., Carbone, L., Cavalleri, A., Cesarini, A., Ciani, G., Congedo, G., Cruise, A. M., Danzmann, K., de Deus Silva, M., De Rosa, R., . . . Zweifel, P. (2016). Sub-femto-g free fall for space-based gravitational wave observatories: Lisa pathfinder results. *Phys. Rev. Lett.*, *116*, 231101. https://doi.org/10.1103/PhysRevLett.116.231101

Bibliography (cont.)

- Armano, M., Audley, H., Baird, J., Binetruy, P., Born, M., Bortoluzzi, D., Castelli, E., Cavalleri, A., Cesarini, A., Cruise, A. M., Danzmann, K., de Deus Silva, M., Diepholz, I., Dixon, G., Dolesi, R., Ferraioli, L., Ferroni, V., Fitzsimons, E. D., Freschi, M., . . . Zweifel, P. (2018). Beyond the required lisa free-fall performance: New lisa pathfinder results down to 20 μHz. *Phys. Rev. Lett.*, *120*, 061101. https://doi.org/10.1103/PhysRevLett.120.061101

- Arun, K. G., Belgacem, E., Benkel, R., Bernard, L., Berti, E., Bertone, G., Besancon, M., Blas, D., Böhmer, C. G., Brito, R., Calcagni, G., Cardenas-Avendaño, A., Clough, K., Crisostomi, M., Luca, V. D., Doneva, D., Escoffier, S., Ezquiaga, J. M., Ferreira, P. G., . . . Zumalacárregui, M. (2022). New horizons for fundamental physics with LISA. *Living Reviews in Relativity*, *25*(1). https://doi.org/10.1007/s41114-022-00036-9

- Auclair, P., Bacon, D., Baker, T., Barreiro, T., Bartolo, N., Belgacem, E., Bellomo, N., Ben-Dayan, I., Bertacca, D., Besancon, M., Blanco-Pillado, J. J., Blas, D., Boileau, G., Calcagni, G., Caldwell, R., Caprini, C., Carbone, C., Chang, C.-F., Chen, H.-Y., . . . Zhdanov, V. I. (2022). Cosmology with the laser interferometer space antenna.

- Babak. (2021). Computational cost and storage for l1,l2, l3 production. *APC.* https://atrium.in2p3.fr/e9dcdf9d-0ee0-41b3-bcb1-8c756cb07ffb

- Babak, S., Gair, J. R. & Porter, E. K. (2009). An algorithm for the detection of extreme mass ratio inspirals in LISA data. *Classical and Quantum Gravity*, *26*(13), 135004. https://doi.org/10.1088/0264-9381/26/13/135004

Bibliography (cont.)

- Baghi, Q., Thorpe, J. I., Slutsky, J., Baker, J., Canton, T. D., Korsakova, N. & Karnesis, N. (2019). Gravitational-wave parameter estimation with gaps in LISA: A bayesian data augmentation method. *Physical Review D*, *100*(2). https://doi.org/10.1103/physrevd.100.022003
- Bartolo, N., Bertacca, D., Caldwell, R., Contaldi, C. R., Cusin, G., De Luca, V., Dimastrogiovanni, E., Fasiello, M., Figueroa, D. G., Franciolini, G., Jenkins, A. C., Peloso, M., Pieroni, M., Renzini, A., Ricciardone, A., Riotto, A., Sakellariadou, M., Sorbo, L., Tasinato, G., . . . Kuroyanagi, S. (2022). Probing anisotropies of the stochastic gravitational wave background with lisa. https://doi.org/10.48550/ARXIV.2201.08782
- Bayle, J.-B., Baghi, Q., Renzini, A. & Le Jeune, M. (2022). *Lisa gw response* (Version 1.1). Zenodo. https://doi.org/10.5281/zenodo.6423436
- Bayle, J.-B., Castelli, E. & Korsakova, N. (2022). *Lisa glitch* (Version 1.1). Zenodo. https://doi.org/10.5281/zenodo.6452904
- Bayle, J.-B., Hartwig, O., Petiteau, A. & Lilley, M. (2022). *Lisanode* (Version 1.4). Zenodo. https://doi.org/10.5281/zenodo.6461078
- Bayle, J.-B., Hees, A., Lilley, M. & Le Poncin-Lafitte, C. (2022). *Lisa orbits* (Version 2.0). Zenodo. https://doi.org/10.5281/zenodo.6412992

Bibliography (cont.)

- Biscoveanu, S., Talbot, C., Thrane, E. & Smith, R. (2020). Measuring the primordial gravitational-wave background in the presence of astrophysical foregrounds. *Phys. Rev. Lett.*, *125*, 241101. https://doi.org/10.1103/PhysRevLett.125.241101
- Consortium, T. e., : Seoane, P. A., Aoudia, S., Audley, H., Auger, G., Babak, S., Baker, J., Barausse, E., Barke, S., Bassan, M., Beckmann, V., Benacquista, M., Bender, P. L., Berti, E., Binétruy, P., Bogenstahl, J., Bonvin, C., Bortoluzzi, D., . . . Zweifel, P. (2013). The gravitational universe. https://doi.org/10.48550/ARXIV.1305.5720
- Cornish, N. J. (2022). Low latency detection of massive black hole binaries. *Physical Review D*, *105*(4). https://doi.org/10.1103/physrevd.105.044007
- Cornish, N. J. & Crowder, J. (2005). LISA data analysis using markov chain monte carlo methods. *Physical Review D*, *72*(4). https://doi.org/10.1103/physrevd.72.043005
- Cornish, N. J. & Littenberg, T. B. (2007). Tests of Bayesian model selection techniques for gravitational wave astronomy. *prd*, *76*(8), Article 083006, 083006. https://doi.org/10.1103/PhysRevD.76.083006
- Cornish, N. J. & Shuman, K. (2020). Black hole hunting with LISA. *Physical Review D*, *101*(12). https://doi.org/10.1103/physrevd.101.124008
- Davis, D., Littenberg, T. B., Romero-Shaw, I. M., Millhouse, M., McIver, J., Di Renzo, F. & Ashton, G. (2022). Subtracting glitches from gravitational-wave detector data during the third observing run.

Bibliography (cont.)

- Dey, K., Karnesis, N., Toubiana, A., Barausse, E., Korsakova, N., Baghi, Q. & Basak, S. (2021). Effect of data gaps on the detectability and parameter estimation of massive black hole binaries with LISA. *Physical Review D*, *104*(4). `https://doi.org/10.1103/physrevd.104.044035`
- Flauger, R., Karnesis, N., Nardini, G., Pieroni, M., Ricciardone, A. & Torrado, J. (2021). Improved reconstruction of a stochastic gravitational wave background with LISA. *Journal of Cosmology and Astroparticle Physics*, *2021*(01), 059–059. `https://doi.org/10.1088/1475-7516/2021/01/059`
- Ford, K. E. S., Bartos, I., McKernan, B., Haiman, Z., Corsi, A., Keivani, A., Marka, S., Perna, R., Graham, M., Ross, N. P., Stern, D., Bellovary, J., Berti, E., O'Dowd, M., Lyra, W., Mac Low, M.-M. & Marka, Z. (2019). Agn (and other) astrophysics with gravitational wave events. `https://doi.org/10.48550/ARXIV.1903.09529`
- Glampedakis, K. (2005). Extreme mass ratio inspirals: LISA's unique probe of black hole gravity. *Classical and Quantum Gravity*, *22*(15), S605–S659. `https://doi.org/10.1088/0264-9381/22/15/004`
- Halloin. (2021). Introduction à lisa. *APC*. `https://indico.in2p3.fr/event/22776`
- Hartwig, O., Bayle, J.-B., Staab, M., Hees, A., Lilley, M. & Wolf, P. (2022). Time-delay interferometry without clock synchronization. *Physical Review D*, *105*(12). `https://doi.org/10.1103/physrevd.105.122008`

Bibliography (cont.)

- Hourihane, S., Chatziioannou, K., Wijngaarden, M., Davis, D., Littenberg, T. & Cornish, N. (2022). Accurate modeling and mitigation of overlapping signals and glitches in gravitational-wave data. *Phys. Rev. D*, *106*(4), 042006. https://doi.org/10.1103/PhysRevD.106.042006
- Jacobs, A., Dunne, J., Moore, C. & Clauset, A. (2015). Untangling the roles of parasites in food webs with generative network models.
- Jaranowski, P. & Królak, A. (2007). Gravitational-wave data analysis. formalism and sample applications: The gaussian case. https://doi.org/10.48550/ARXIV.0711.1115
- Katz, M. (2021). *Mikekatz04/bbhx: First official public release* (Version v1.0.0). Zenodo. https://doi.org/10.5281/zenodo.5730688
- Katz, M. L. (2022). *Mikekatz04/gbgpu: First official public release!* (Version v1.0.0). Zenodo. https://doi.org/10.5281/zenodo.6500434
- Katz, M. L., Chua, A. J., Speri, L., Warburton, N. & Hughes, S. A. (2021). Fast extreme-mass-ratio-inspiral waveforms: New tools for millihertz gravitational-wave data analysis. *Physical Review D*, *104*(6). https://doi.org/10.1103/physrevd.104.064047
- Klein, A., Barausse, E., Sesana, A., Petiteau, A., Berti, E., Babak, S., Gair, J., Aoudia, S., Hinder, I., Ohme, F. & Wardell, B. (2016). Science with the space-based interferometer eLISA: Supermassive black hole binaries. *Physical Review D*, *93*(2). https://doi.org/10.1103/physrevd.93.024003

Bibliography (cont.)

- Korol, V., Toonen, S., Klein, A., Belokurov, V., Vincenzo, F., Buscicchio, R., Gerosa, D., Moore, C. J., Roebber, E., Rossi, E. M. & Vecchio, A. (2020). Populations of double white dwarfs in Milky Way satellites and their detectability with LISA. *aap*, *638*, Article A153, A153. https://doi.org/10.1051/0004-6361/202037764
- Korol, V., Toonen, S., Klein, A., Belokurov, V., Vincenzo, F., Buscicchio, R., Gerosa, D., Moore, C. J., Roebber, E., Rossi, E. M. & Vecchio, A. (2020). Populations of double white dwarfs in milky way satellites and their detectability with lisa. *A&A*, *638*, A153. https://doi.org/10.1051/0004-6361/202037764
- Laghi, D., Tamanini, N., Del Pozzo, W., Sesana, A., Gair, J., Babak, S. & Izquierdo-Villalba, D. (2021). Gravitational-wave cosmology with extreme mass-ratio inspirals. *Monthly Notices of the Royal Astronomical Society*, *508*(3), 4512–4531. https://doi.org/10.1093/mnras/stab2741
- Lammers, U., Petiteau, A., Texier, D., McNamara, P. & Jennrich, O. (2019). Lisa science operations assumptions document. *SOC/DPC*. https://atrium.in2p3.fr/479720a5-85c8-4a35-a5a3-d7b168b53bb6
- Lang, R. N. & Hughes, S. A. (2006). Measuring coalescing massive binary black holes with gravitational waves: The impact of spin-induced precession. *prd*, *74*(12), Article 122001, 122001. https://doi.org/10.1103/PhysRevD.74.122001
- Le Jeune, M. & Babak, S. (2022). *Lisa data challenge sangria (ldc2a)* (Version v2). Zenodo. https://doi.org/10.5281/zenodo.7132178

Bibliography (cont.)

- Mangiagli, A., Caprini, C., Volonteri, M., Marsat, S., Vergani, S., Tamanini, N. & Inchauspé, H. (2022). Massive black hole binaries in lisa: Multimessenger prospects and electromagnetic counterparts. https://doi.org/10.48550/ARXIV.2207.10678
- Marsat, S., Baker, J. G. & Canton, T. D. (2021). Exploring the Bayesian parameter estimation of binary black holes with LISA. *prd*, *103*(8), Article 083011, 083011. https://doi.org/10.1103/PhysRevD.103.083011
- Nelemans, G., Yungelson, L. R. & Portegies Zwart, S. F. (2004). Short-period AM CVn systems as optical, X-ray and gravitational-wave sources. *mnras*, *349*(1), 181–192. https://doi.org/10.1111/j.1365-2966.2004.07479.x
- Paczkowski, S., Giusteri, R., Hewitson, M., Karnesis, N., Fitzsimons, E. D., Wanner, G. & Heinzel, G. (2022). Postprocessing subtraction of tilt-to-length noise in LISA. *prd*, *106*(4), Article 042005, 042005. https://doi.org/10.1103/PhysRevD.106.042005
- Pozzo, W. D., Sesana, A. & Klein, A. (2018). Stellar binary black holes in the LISA band: A new class of standard sirens. *Monthly Notices of the Royal Astronomical Society*, *475*(3), 3485–3492. https://doi.org/10.1093/mnras/sty057
- Prince, T. A., Tinto, M., Larson, S. L. & Armstrong, J. W. (2002). LISA optimal sensitivity. *prd*, *66*(12), Article 122002, 122002. https://doi.org/10.1103/PhysRevD.66.122002
- Roy, V. (2019). Convergence diagnostics for markov chain monte carlo. https://doi.org/10.48550/ARXIV.1909.11827

Bibliography (cont.)

- Savalle, E., Gair, J., Speri, L. & Babak, S. (2022). Assessing the impact of instrumental calibration uncertainty on LISA science. *prd*, *106*(2), Article 022003, 022003. https://doi.org/10.1103/PhysRevD.106.022003
- SAVALLE, E., LEJEUNE, M. & BABAK, S. (2022). *Lisa multifisher*. Zenodo. https://doi.org/10.5281/zenodo.7322980
- Staab, M., Bayle, J.-B. & Hartwig, O. (2022). *Pytdi* (Version 1.2.1). Zenodo. https://doi.org/10.5281/zenodo.6867012
- Tamanini, N., Caprini, C., Barausse, E., Sesana, A., Klein, A. & Petiteau, A. (2016). Science with the space-based interferometer eLISA. III: Probing the expansion of the universe using gravitational wave standard sirens. *Journal of Cosmology and Astroparticle Physics*, *2016*(04), 002–002. https://doi.org/10.1088/1475-7516/2016/04/002
- Wang, Y., Shang, Y. & Babak, S. (2012). Extreme mass ratio inspiral data analysis with a phenomenological waveform. *Physical Review D*, *86*(10). https://doi.org/10.1103/physrevd.86.104050
- Xuan, J., Lu, J. & Zhang, G. (2019). A survey on bayesian nonparametric learning. *ACM Computing Surveys (CSUR)*, *52*, 1–36.