LISA Data Analysis Work Packages

N/Ref :	LISA-LCST-SGS-WPD-001
Title	**LISA Data Analysis Work Packages**
Abstract	LISA DPC Definition and main contributions

	Name	**Date**	**Signature**
Prepared by	LISA WP writing team	2018/01/19	
Checked by			
Checked by (QA)			
Approved by			

Document Change Record

Contributor List

Author's name	Institute	Location
Babak Stanislav	APC	Paris, France
Barausse Enrico	IAP	Paris, France
Barack Leor	Univ. of Southampton	Southampton, UK
Caprini Chiara	APC	Paris, France
Christensen Nelson	ARTEMIS	Nice, France
Colpi Monica	Univ. Milano Bicocca	Milano, Italy
Cornish Neil	Montana State University	Bozeman, MT, US
Ferraioli Luigi	ETH	Zurich, Switzerland
Gair Jonathan	Univ. of Edinburgh	Edinburgh, UK
Larson Shane	Northwestern University	Chicago, IL, US
Littenberg Tyson	NASA-Marshall	Huntsville, AL, US
Hewitson Martin	AEI	Hannover-Germany
McWilliams Sean	West Virginia University	Morgantown, WV, US
Nardini Germano	Univ. of Bern	Bern, Switzerland
Petiteau Antoine	APC	Paris-France
Rossi Elena Maria	Leiden Observatory	Leiden, Netherlands
Sesana Alberto	Univ. of Birmingham	Birmingham, UK
Sweta Shah	AEI	Hannover-Germany
Tamanini Nicola	CEA	Saclay, France
Vallisneri Michele	Caltech/NASA-JPL	Pasadena, CA, US
Vecchio Alberto	Univ. of Birmingham	Birmingham, UK

Ver.	Date	Author	Description	Pages
0.1	2017-09-13	Edinburgh team	Initial version	all
0.3	2017-10-27	LISA WP writing team	preliminary draft before the first telecon	all
0.5	2017-10-29	LISA WP writing team	first complete draft after the final telecon	all
0.6	2018/01/19	WP writing team (LISA)	Current Version	

Distribution list

Recipient	Restricted	Not restricted
LISA Consortium	✗	

Contents

Purpose and Scope

The purpose of this document is to describe the work needed to construct a data analysis infrastructure for the analysis of LISA data that will allow us to meet all of the key science objectives. This list was started at a meeting held in Edinburgh on September 11th to 13th 2017. At this meeting, each of the science objectives identified in the LISA proposal were considered, and the work required to meet the objectives was discussed and then divided into packages, that were subsequently grouped when common elements were identified. This document describes the current list of work packages identified in this way. This list is designed to form a basis for data analysis planning, but is not expected to remain unchanged between now and launch. Additional work packages will be added as additional needs are identified and other packages will be removed as the work in them is done or they are made obsolete by changes or developments elsewhere.

The work packages are grouped into eight groups according to their theme — group 1 concerns waveform modelling, group 2 covers data analysis tools that are common to multiple pipelines, group 3 covers instrument response modelling, group 4 concerns low-latency analysis pipelines, group 5 concerns codes for identification of sources, group 6 covers source catalogue construction, group 7 covers multi-band analysis tools and finally group 8 covers the tools required to allow final delivery of key science objectives of the mission.

For each work package within each group, a priority has been assigned according to the following convention

0 = urgent (couple of months, needed in phase A)

1 = 2 years (end of phase A)

2 = 5 years (demonstration of capabilities at TRL6)

3 = 10 years (development of sci ops)

At the start of each section, a table summarises the priorities of all work packages within the group described in that section.

1 Waveform modelling

The existence of accurate waveform models is critical to extract maximum science from LISA observations. Accurate astrophysical parameter estimates and precision tests of general relativity will only be possible if we have models of gravitational wave signals in general relativity that are accurate to within a fraction of a cycle of phase over the whole LISA observation. LISA places specific requirements on waveform models that are different to those required by LIGO. Signals are typically much longer and signal-to-noise ratios are higher. While there is a very active waveform modelling community, partially driven by the desire to have models ready for LIGO observations, and partly by intrinsic interest in the two-body problem in GR, this community may not provide exactly what LISA requires unless there is specific direction from the LISA community. Waveform modelling is therefore an important element of the development of LISA data processing pipelines and the work packages in this section are focussed on that.

There is one high priority work package, 1.1, which exists to put the other work packages in context. This work package will quantify what will be needed in terms of the physics included in waveform models to ensure LISA can deliver its science, and assess the time and manpower requirements to achieve that. The remaining work packages, 1.2-1.8, aim to oversee the development of accurate and faithful general-relativistic waveform models of potential sources, design and implement tools for fast waveform generation, and develop standard-format interfaces and visualization tools. Different classes of sources present distinct modelling challenges and are therefore dealt with under separate WPs (1.2-1.7), with WP 1.8 covering common waveform interfaces and tools. Accurate waveform models underpin many of the detection strategies, and are crucial for the interpretation of detected sources, extraction of source parameters, and removal of foreground signals. As an ideal standard, we require that models are sufficiently accurate that "systematic" modelling error is a subdominant component of the total noise budget; this standard has essentially been achieved for some sources but would require much more work for others. Work is also needed in assessing the likelihood and importance of various sources of astrophysical perturbations, and these need to be accommodated in the models if necessary. Finally, it is desirable to design waveform families that incorporate parameterized deviations from GR, or the effect of exotic physics within GR.

WP	Description	Priority
1.1	Assess LISA waveform modelling requirements	1
1.2	Provide EMRI waveforms	2
1.3	Provide MBHB waveforms	2
1.4	Provide GB waveforms	2
1.5	Provide IMRI waveforms	3
1.6	Provide SOBHB waveforms	2
1.7	Provide waveforms for cosmic strings and other modelled transient events	3
1.8	Waveform interface and tools	1

Table 1: Work packages on Waveform modelling.

1.1 Assess LISA waveform modelling requirements

Goals/Motivation

- Identify the necessary physics that must be included in waveform models of each type in order to deliver LISA science goals.
- Assess the effort required to build waveforms of the necessary complexity, in terms of time, computational cost and manpower.
- This work package provides input for WP 1.1-1.7, by identifying at what stage prior to launch the LISA consortium needs to actively develop (or push the community to develop) waveform models of the complexity required for LISA data processing.

Outputs

- White paper describing current status of EMRI waveform models, and physical effects that are missing, with reference to LISA key-science deliverables in WP group 8.
- White paper with a similar description for MBHBs.
- Plan for development of EMRI models of necessary complexity, with estimated manpower required.
- Similar plan for MBHBs.
- Assessment of impact on key-science if waveform models are not up to required standard by launch.

Time-frame

- Assessment needed by end of phase A, so that manpower/funding requirements can be evaluated.
- Continual review and assessment needed up until launch, to react to developments in the waveform modelling community.

Possible sub work-packages

- Modelling requirements for EMRIs.
- Modelling requirements for MBHBs.

Dependencies

- Feeds directly into work packages 1.2-1.7.
- Strongly influenced by WP group 8. Key science deliverables place requirements on waveforms that must be included in the assessment here.

1.2 Provide EMRI waveforms

Goals/Motivation

- Provide an accurate model of EMRI waveforms for a BH CO, within GR, across the astrophysically relevant parameter space. The model should allow generic MBH and CO spin magnitudes and orientations, generic orbital inclination, and generic eccentricity in the relevant range. It should be phase-accurate to within a fraction of a radian over the entire in-band portion of the inspiral.
- For non-BH COs, quantify the effect of CO internal structure (mass and mass-current quadrupoles and higher moments) on the EMRI waveform. If deemed important, devise a model of non-BH EMRIs to include relevant internal-structure effects.
- Assess probability and impact of environmental perturbations (gas, simultaneous EMRIs, other perturbers).
- Design and implement a framework for incorporating self-force-based numerical calculations, as they become available, into a flexible semi-analytical Kludge model that enables fast production of waveform templates.
- Revise estimates of EMRIs as a source of stochastic confusion noise in light of astrophysical EMRI population models as they improve.

Outputs

- Method and code (+ publications reporting these) for calculating an EMRI waveform for any given set of parameters in the relevant range.
- Method and code (+ publications reporting these) for producing Kludge waveforms informed by accurate self-force-based calculations.
- Publications reporting results from studies of CO structure and environmental effects; method and code for implementing these effects if they are likely to introduce significant systematic bias.

Time-frame

- Until L-5yr: Routinely monitor progress of modelling effort in relation to above goals, and seek to incorporate results as they become available to improve Kludge models for DA studies.
- Around L-5yr: Conduct a review of waveform readiness and assess remaining tasks. If insufficient progress, promote to higher priority, and consider allocating resources from within the project for a targeted modelling effort before launch.
- During mission: support the analysis of candidate EMRI signals with high-accuracy waveform calculations at narrowed-down regions of the parameter space.

Possible sub work-packages

- Effects of CO internal structure
- Environmental effects
- Resonances
- smoking-gun features of deviation from GR
- Kludge models

Dependencies

- Feeds directly into 1.8 (waveform interface and tools) and (indirectly) into 5.4 (EMRI detection and parameter extraction).
- Feeds into 1.5 (IMRI waveforms) by providing calibration and test for IMRI models.
- Supplies test and training data for 4.1 (low-latency pipeline)
- Feeds into 8.7 (tests of GR)
- Depends on WP 6, WP 3.

1.3 Provide MBHB waveforms

Massive black hole binaries are expected to be the strongest source in the LISA band. The full signal contains three distinguishable parts: the *inspiral* (adiabatic evolution through loss of angular momentum and energy via gravitational radiation), the *merger* (highly relativistic regime when a single deformed object is formed), and the *ringdown* (the single remnant MBH relaxes to Kerr via emission of ring-down gravitational radiation). Depending on the total mass of the system we can expect to detect some or all parts of the GW signal. Some MBHBs will be detected with extremely high SNRs, necessitating a high-precision model.

Goals/Motivation

- Develope/improve a model of MBHB waveforms that accounts for arbitrary spin configurations and relevant ranges of eccentricities and mass ratios. This requires the combination of an accurate post-Newtonian model for the inspiral part, with full numerical simulations for the merger and ringdown.
- The desirable accuracy of the model should be such that the systematic error is below the statstical ($\sim$1/SNR).

- Fast waveform generation is essential for data analysis, motivating the development of fast frequency-domain waveform generators.
- Design an extension of the IMR (inspiral-merger-ringdown) waveforms beyond GR either in a phenomenological (parametrized) way or for specific alternative models of gravity. Also include possible MBH mimickers within GR (like Boson stars).

Outputs

- Catalogue of numerical waveforms spanning the expected parameter space in mass ratio, spins (magnitude and orientation) and eccentricity.
- Fast generator of IMR waveforms, cross-validated against avaliable NR waveforms, with an extensive systematic study across the parameter space.
- Systematic study of the impact of subdominant modes, eccentricity and precession on detection and parameter estimation.
- Parametrized extension of IMR waveforms beyond GR: propagation effects, deviations in the ringdown, deviatons in the inspiral part, polarizations.

Time-frame

- Some IMR waveform models already exist and can be used in upcoming LISA data challenges.
- Until L-5yr, Improve IMR waveforms by including precession and eccentricity.
- Until L-5yr: Review of the status of waveforms with parametrized deviations from GR, and promote to higher priority if needed.
- Around L-5yr. Review existing models and promote to higher priority if accuracy is not sufficient or some key features are missing. Find and allocate resources to complete the models before L-2yr.

Possible sub work-packages

- Generation of numerical waveforms across the expected parameter space in mass ratio, spins and eccentricity
- Fast waveform generation (e.g., frequency domain)
- Study of systematic errors across parameter space.
- Extension to include possible non-GR effects or exotic physics within GR.

Dependencies

- Feeds into/Overlaps with WP1.8
- Depends on WP3
- Overlaps strongly with WP4
- Mild dependence on WP6

1.4 Provide GB waveforms

Compact galactic binaries, composed of compact stellar remnants—White-Dwarfs, Neutron Stars or Black Holes—are expected to be among the most numerous and loudest signals detected by LISA. Population models indicated that WD-WD binaries will be so numerous that the combined signal from unresolved systems will be a limiting "noise" source across part of the LISA band. Galactic binaries orbit in the weak field regime and can be accurately modeled using post-Newtonian theory, though astrophysical effects such as mass transfer, tides and three body dynamics can add complexity to the signals.

Goals/Motivation

- Provide accurate and waveforms that include all astrophysical and relativistic effects
- Fast waveform models are needed since we expect to detect tens of thousands of systems
- Sky localization for electromagnetic counterpart detection is possible for higher-frequency systems

Outputs

- Low-order post-Newtonian waveforms, up to second order in the frequency derivative, for detached WD/NS/BH binaries
- Low-order post-Newtonian waveforms for mass transferring WD binaries
- Low-order post-Newtonian waveforms including the effects of additional bodies (e.g. hierarchical triples)
- Low-order post-Newtonian waveforms including tidal interactions and resonances
- Fast frequency-domain and wavelet-domain versions of all GB waveforms folded with the instrument response

Time-frame

- Fast frequency-domain waveforms for detached and interacting systems, including a simplified instrument response already exist
- Waveforms for GBs in hierarchical triples already exist. Fast frequency domain versions needed. Months of work.
- Tidal effects have been studied, but have not been incorporated in the waveform modelling. Anticipate one year to complete
- Work on incorporating more realistic instrument response in fast frequency-domain waveforms is underway. Year of work.

Possible sub work-packages

- Frequency domain instrument response
- Tidal effects on orbits
- Fast wavelet domain waveforms, possible via a ROQ

Dependencies

- Fast waveform generation tools (WP 1.8)
- Instrument response modeling (WP 3)
- Source catalogues (WP 6)

1.5 Provide IMRI waveforms

Goals/Motivation

- Provide an accurate model of IMRI waveforms, within GR, for both MBH-IMBH systems and IMBH-CO systems where the CO is a stellar-mass BH. The model should allow generic spin configuration, orbital inclination and eccentricity in the relevant range. It should be phase-accurate to within a fraction of a radian over the entire in-band portion of the inspiral.
- For IMBH-CO IMRIs where the CO is not a BH, quantify the effect of CO internal structure (mass and mass-current quadrupoles and higher moments) on the IMRI waveform. If deemed important, devise a model of non-BH IMRIs to include relevant internal-structure effects.
- For both above items, assess reliability of "perturbative" waveforms (based on self-force calculations alone) through comparison with exact NR waveforms. Determine whether the former can suffice for data analysis.
- Assess probability and impact of environmental perturbations (gas, simultaneous captures, other perturbers).
- Design and implement a framework for incorporating numerical IMRI waveform calculations, as they become available, into a flexible semi-analytical Kludge model that enables fast production of waveform templates.

Outputs

- Method and code (+ publications reporting these) for calculating an IMRI waveform for any given set of parameters in the relevant range.
- Method and code (+ publications reporting these) for producing Kludge IMRI waveforms.
- Publications reporting results from studies of CO structure and environmental effects; method and code for implementing these effects if they are likely to introduce significant systematic bias.

Time-frame

- Until L-5yr: Routinely monitor progress of modelling effort in relation to above goals, and seek to incorporate results as they become available to improve Kludge models for DA studies.
- Around L-5yr: Conduct a review of waveform readiness and assess remaining tasks. If insufficient progress, promote to higher priority, and consider allocating resources from within the project for a targeted modelling effort before launch.
- During mission: support the analysis of candidate IMRI signals with high-accuracy waveform calculations at narrowed-down regions of the parameter space.

Possible sub work-packages

- NR methods for large mass ratios and interface with self-force
- Effects of CO internal structure
- Environmental effects
- smoking-gun features of deviation from GR
- Kludge models

Dependencies

- Feeds from 1.2 (EMRI waveforms) for test and calibration.
- Feeds directly into (and overlaps with) 1.8 (waveform interface and tools) and 5.6 (IMRI detection and parameter extraction).
- Feeds into 8.7 (tests of GR)
- Depends on WP 6, WP 3

1.6 Provide SOBHB waveforms

Stellar-mass BH binaries merging in the LIGO/Virgo band may be observable by LISA in their inspiral phase. Merger rates in the LIGO/Virgo band suggest tens/hundreds of such systems may be potentially detectable. However, sufficiently accurate waveforms are necessary to disentangle these sources from the noise, and to correctly estimate their parameters (particularly merger times and sky localization, which are crucial to alert ground interferometers, to test GR, and to assess the presence of possible electromagnetic counterparts).

Goals/Motivation
Provide accurate and faithful SOBHB waveforms, to enable

- the detection of SOBHBs before merger, and alert ground interferometers in advance;
- the retroactive identification of weak SOBHB inspiral signals in LISA data, based on merger signals detected by LIGO/Virgo;
- unbiased estimates of source parameters, so as to allow for multi-band astrophysical population analyses, tests of GR (e.g. vacuum dipole radiation) and prompt sky localization for electromagnetic searches.

Outputs

- Analysis to determine the highest PN orders (in phase and amplitude) needed to provide unbiased parameter estimation for the expected astrophysical population of SOBHB, and provide faithful GR waveforms.
- Method and code for producing long faithful precessing-spin inspiral-merger-ringdown GR waveforms for cross-band detection and parameter estimation.
- Method and code for producing parametrized waveforms allowing for deviations away from GR and for possible interactions with matter/massive perturbers.

Time-frame

- This WP is lower priority than other modelled sources, because SOBHBs are low-SNR sources whose subtraction is not likely to be an issue for the detection and parameter estimation of other sources. Most of the necessary tools have already been developed: We envisage a preliminary pipeline in 3 years and the final pipeline in 5 years.

Possible sub work-packages

- Faithful inspiral PN waveforms.
- Faithul inspiral-merger-ringdown waveforms.
- Faithful waveforms with parameterized deviations from GR/environmental effects.

Dependencies

- MBHB waveforms (WP 1.3)
- Source catalogues (WP 6).
- Framework for common data analysis tools (WP 2.1, WP 2.2, WP 2.3).
- Collaboration with population synthesis modellers.
- Collaboration with LIGO/Virgo.

1.7 Provide waveforms for cosmic strings and other modelled transient events

Goals/Motivation

- Implement waveform models for modelled short-duration transient events, e.g., cosmic string cusps and kinks, extreme-mass-ratio bursts, or other hyperbolic black hole encounters.
- Models exist for gravitational waves from cosmic strings and hyperbolic encounters, but more work is needed for extreme-mass-ratio bursts.

Outputs

- Tools to generate waveforms for each source type.

Time-frame

- Preliminary models exist and have been used in MLDCs (cosmic string cusps). LDCs with transient signals will be issued in the future, but existing frameworks likely sufficient.
- Final models are low priority due to low event rate and relative simplicity of transient searches. Needed L-2y.
- If other transient sources are predicted between now and L-2y, additional models should be added.

Possible sub work-packages

- Waveform models for cosmic string cusps and kinks.
- Waveform models for extreme-mass-ratio bursts.

Dependencies

- EMRI waveform modelling (WP 1.2).

1.8 Waveform interface and tools

Waveforms developed either in the wider scientific community or within dedicated WPs should adhere to certain standards so that they can be directly used in the data analysis. This requires the design of well-defined interfaces and tools.

Goals/Motivation

- Interfaces: A particular waveform model should be parametrized in a way suitable for the data analysis (e.g., seeking to reduce parameter correlation and allowing running MCMC)
- Interfaces: we need to make sure that the format (programming language and output) is suitable for applying the response function (this includes phase and amplitude modulation due to detector motion and TDIs)
- Interface: we need an interface to read astrophysical catalogues of sources (with possible downselection) and pass parameters to the waveform generation tools.
- Tools: we need visualization tools as a part of the waveform testing and for presentation purposes.
- Tools: We need to validate the produced waveforms.
- Tools: We need to optimize the waveforms in terms of computational efficiency.

Outputs

- An interface that convert between modelling and data-analysis parametrizations, and can inform the choice of parameters for waveform catalogues.
- An Interface between astrophysical catalogues and waveform generation
- A tool (either generic or source-specific) that applies the response function to the waveforms (which are typically generated in the source frame)
- Visualization tools for waveforms in the time/frequency/time-frequency domains. This includes raw polarization, waveforms including response, and characteristic strain (averaged over the sky, polarization and inclination)
- Tools for computing SNR, overlaps, likelihood and sanity checks (as unit tests)
- Optimization tools. For slow waveforms we need to create one of the following approximations: phenomenological template-based, reduced order model (ROM) and/or reduced order quadrature (ROQ) tool for likelihood computations.

Time-frame

- Interfaces: should be available now. Will be realized and evolve within LISA data challenges
- Tools: Visualization and validation tools should be available within 1 year.

- Optimization tools: Ongoing long-running project. The common infrastructure for ROM/ROQ should be available within a year. ROM/ROQ for particular models should be developed as we need within half a year. The phenomenological models (if not developed within the waveform working packages) should be developed as we need within half a year time frame.

Possible sub work-packages

- WP on the interfaces (see above)
- WP on the visualization and validation of the waveforms (see above)
- WP on the optimization of the waveforms (see above)

Dependencies

- Depends on the WP1.2-1.7, WP3.1-3.3, WP5.1-5.7, WP6.1-6.5

2 Data analysis tools

This section highlights a number of work packages associated with data analysis tools in general, and in particular, those which form a common toolkit that can be used to assemble pipelines, produce higher level data products, analysis instrument performance, and extract scientific conclusions from LISA data. Wherever possible, effort should be made to produce common tools with broad enough interfaces to allow them to be used in multiple contexts, thus minimising duplication and easing maintenance and testing. Table 2 summarises the work packages of this group.

WP	Description	Priority
2.1	Data analysis framework	2
2.2	Data analysis library	2
2.3	Define data exchange formats (including waveforms)	1
2.4	Alert distribution tools	3
2.5	Tools for detecting unmodelled signals	2
2.6	Tools to deal with delensing	3
2.7	Tools for modelling instrument performance	0
2.8	Tools for assessing instrument/mission design choices (performance, impact on science, etc)	0
2.9	Tools for generating and managing data quality indicators	2
2.10	Tools to assess impact of gaps, stationarity, data rates, spectral lines, glitches, etc on science output	0

Table 2: Work packages on data analysis tools.

2.1 Data analysis framework

Goals/Motivation In order to build a consistent and modular data analysis pipeline, all the processing codes have to be implemented in a common framework. This framework should also be the same than the one used for simulation and for preprocessing (L0 $\rightarrow$ L1) tasks, given their high level of interaction with the scientific data analysis. The aim of this framework is to ease the interoperability between the various data processing steps, and ease the execution of the pipeline on arbitrary infrastructure. While its precise functionalities have to be defined in the design phases of the mission along with the DPC, some standard tools can be provided to the community at early stage to start building a collaborative environment.

This workpackage thus has two objectives:

- Provide a developer toolkit which can host the common tools developed by the DAWG (and the simulation WG). Experience acquired from LISAPF analysis environment (LTPDA) could be exploited to target an interactive environment with high level facilities.
- Gather and synthetize the important features that the DPC will have to provide to the consortium on this aspect.

Outputs Some tools have already be made available within the so called proto-DPC environment, which includes control version system, continuous integration and containers. The expected outputs are thus an improved version of the existing proto-DPC, following on requests and needs addressed by the team.

The second main outputs of this workpackage is the definition of the requirements of the future framework provided by the DPC, that could take the form of a "Data analysis framework requirement" document.

Time-frame The development of the proto-DPC will continue over the next design phases (A → B1).

The "Data analysis framework requirement" document is due for end of phase B1.

Possible sub work-packages

- Development environment
- Computing environment
- Interface and vizualization

Dependencies The framework is desirable for most of the WPs since it will ease the development of codes and provide the guideline and the tool to directly integrate the codes in the DPC pipelines.

2.2 Data analysis library

Goals/Motivation The aim of this workpackage is to identify, design and implement common and key data analysis algorithms and tools such as Fourier series analysis tools, filtering tools, sampling, optimization and vizualization tools. A common reference on those low or intermediate level libraries would avoid duplication of efforts to develop them, and ease the comparison of the results obtained by different algorithms at higher level. In order to fullfill its objectives, the library should be modular, easy-to-use, and should have good performances.

A unique definition and implementation of the constants should also be provided by this workpackage.

Outputs The expected outputs of this workpackage is twofolds:

- Provide and maintain a list of standard existing tools (with version reference number) that will constitute the lowest level of the data analysis library (for example FFTW3.X). This output could take the form of a common container, or feed the data analysis framework (see WP2.1)
- Design and implement the LISA data analysis library. The design should use an appropriate format which will encourage discussions and ease its refinement at the consortium level (ie be explicit and understood by non specialist).

Time-frame A first version should be put in place along the first LDC. A validated version should be delivered for the end of phase A, early in B1 at the latest. The validation process should focus on the properties listed above (performances, modularity, ...), rather than completeness. The development and maintenance of the data analysis library will continue all along the mission.

Possible sub work-packages

- Library containing basic tools: Fast Fourier Transform, filter, Power Spectral Density
- Samplers and optimisation tools
- Vizualisation tools adapted to the data format or catalogue format

Dependencies

- Strong dependencies with WP 2.1 since this tools have to be developed and integrated to the framework
- Dependencies to the data model have to be minimized or synthetised in a simple interface. On the other hand, we expect strong interaction of this workpackage with WP 2.3 on the design of the data model regarding performances measurement.

2.3 Define data exchange formats (including waveform)

Goals/Motivation

- We need to create flexible, efficient, and extensible formats that implement interfaces, between data-analysis pipeline elements, between data-analysis and waveform-generation codes, etc.
- A subsets of formats will be used internally, but others will be integral to LISA data archive and data releases.
- We need to leverage future-proof standards (e.g., HDF5, possibly FITS) that have broad adoption, abundant read–write libraries, guarantee of future support.
- Emphasize inclusion of provenance and tracking information to ensure reproducibility and reliability of results.
- Formats should be supported by a high-level I/O library (or possibly a plug-in to the HDF5 library), in multiple language, so that downstream pipeline components access the files in a uniform, transparent way, supporting updates and extensions.
- Formats should be specified in formal specification documents, but also in schemata or databases that allow the validation of files, and that are used directly to implement parsing in the I/O library.

Outputs

- Library of file formats with specification documents and schemata.
- Data exchange white paper.
- I/O library and example files in multiple languages.

Time-frame

- First prototype version in place during/after first LDC.
- Alpha-quality version at beginning of phase A.
- Beta-quality version at end of Phase A.

Possible sub work-packages See outputs.

Dependencies

- Design must be harmonized and updated as WP 2.1 develops.
- I/O library would be part of WP 2.2.
- Formats would be used by many package components across all WPs. It is essential that we provide mechanisms to relay needs, requests, and feedback from all activities to the format designers and maintainers; and conversely for activities to absorb format updates and tools.

2.4 Alert distribution tools

Goals/Motivation LISA enables the unique opportunity to alert electromagnetic observing partners of SMBH coalescence events well in advance of the merger. As the binaries evolve their source parameters, including sky location, distance and orientation, will be known to increasing precision, with rapid improvement during the last stages of the inspiral. To maximize EM community involvement the procedure for alerting observers to new sources, and predictably updating parameters for known sources, must be standardized, automated, and tested in advance of the mission. LISA will also need to broadly communicate information about short-lived transients that will not necessarily benefit from weeks to months of improving localization, a scenario directly comparable to, and adaptable from, the LIGO/Virgo EM Alert activities.

Outputs

- Standardized format and database for disseminating and storing information about new and updated sources.
- Tools for autonomously creating and distributing alerts and updates to SMBHs.
- Tools for autonomously creating and distributing alerts to short-lived transients
- Tools for autonomously receiving alerts and updates from LISA analyses to be shared with the EM community.
- Scheduler to automatically run pipelines that refine SMBH parameters and distribute updated parameters to community.

Time-frame

- This work must build off of existing infrastructure used in time-domain astronomy (e.g. LVAlert, GCN, etc.). Tools used in time-domain astronomy may undergo a revolution before launch as new facilities (e.g. LSST) come online and LIGO/Virgo transition to open data. The LISA tools must be adaptable to whatever has been adopted by the wider community at the time of launch.

Possible sub work-packages

- Run mock EM alerts as part of LDCs
- Develop tools for autonomously communicating between LISA and the community about new sources (either long-lasting or impulsive) in low-latency and updates to known sources.
- Determine optimal cadence for updating SMBH parameters especially in the late stages before merger.

Dependencies

- Low-latency pipelines 4.1,4.2,4.5

2.5 Tools to deal with delensing

LISA provides a measurement of the luminosity distance of the GW emitting binary that, together with an identification of the redshift of the source (either via an EM counterpart or through statistical techniques), can be used to probe the accelerated expansion of the universe and cosmological parameters. The instrumental error on d_L can be lower than 1% for high SNR sources; however, gravitational lensing from the Large Scale Structure (at intermediate and high redshift) and peculiar velocities (at low redshift) introduce an intrinsic uncertainty in the determination of the true value of d_L. The error due to weak lensing increases with redshift and is of the order 3-5% already at $z = 3$, therefore fully dominating the error budget, for example, of the vast majority of MBHBs. The lensing error can be beaten down statistically, by increasing the number of GW sources. However, developing techniques for removing the uncertainty in d_L by estimating the lensing magnification of GW sources using weak lensing maps is crucial to improve the accuracy with which the cosmological paradigm can be tested.

Goals/Motivation

- Clean as much as possible the measurement of the luminosity distance from the (de)magnification due to lensing from the Large Scale Structure in order to reduce the error with which LISA can constrain cosmological parameters
- Use weak lensing to probe the cosmological paradigm beyond FLRW background evolution: measure the lensing magnification power spectrum and possibly its angular dependence, which contain information on cosmological parameters

Outputs

- Reliable probability distribution functions for the lensing magnification in order to get accurate estimates of lensing effects, ideally at all redshift at which LISA sources are present ($0 < z < 10$)
- Where it is possible at low redshift ($z \lesssim 3$, although the exact threshold is not certain, so determining this is part of the work in this package), use magnification maps reconstructed from the shear and flexion data of large cosmological surveys such as Euclid
- At higher redshifts, use techniques of ray tracing through cosmological N-body simulations to get a prediction of the magnification probability distribution function to fold into astrophysical inference. Investigate the possibility of using CMB lensing to reconstruct lensing maps at high redshift.
- Reconstruct the angular power spectrum of the lensing magnification.

Time-frame

- This WP is low priority and for its development and publications should follow the time frame of WP 8.10 on the estimation of cosmological parameters (5-10y)
- The analyses connected to this WP should be kept up to date in what concerns the measurement of lensing maps by EM cosmological surveys
- The full analysis with data and the paper writing can be carried out on a schedule L+5y

Possible sub work-packages No sub work-packages have yet been identified for this WP.

Dependencies

- WP 8.10 on the use of standard sirens to estimate cosmological parameters
- WP 7.1 and 7.5 for the MoU guaranteeing the access to cosmological surveys data (Euclid)

2.6 Tools for modelling instrument performance

Goals/Motivation

- Tools to assemble a bottom-up Current Best Estimate (CBE) of the observatory performance.
- Tools should allow inclusion of frequency-dependence of each noise source.
- Allow for potential correlations between different sources of noise.
- Include performance of TDI and preprocessing algorithms, and potential correlations produced from individual noise sources.
- Produce easy to use and rapid assessment tools which can be used to track the observatory performance as a function of key parameters and noise levels.
- Output different TDI combinations.
- Assess sky/polarisation averaged SNR of observatory to typical LISA sources. Produce the horizon distances for canonical sources (where applicable)
- Allow for evolution in knowledge of noise sources.
- Tools to assess detection and merger times for MBHBs.
- Tools to assess accuracy of parameter estimation for different observatory configurations/-parameterisations.

Outputs

- A set of tools to produce frequency-domain descriptions of the observatory performance.

Time-frame

- Initial set of tools for primary TDI output required in Phase A.
- More sophisticated assessment of different TDI outputs needed early in B1.
- Tools should continue to evolve as more noise sources and correlations are studied.

Possible sub work-packages

- Correlation effects in TDI.
- Studies of preprocessing and TDI performance, including all realistic effects (clock-noise removal, multiple clocks, common timing grid, etc).
- Waveforms and SNR tools.
- Studies of individual noise sources (or groups of noise sources).

Dependencies No dependencies yet identified.

2.7 Tools for assessing instrument/mission design choices (performance, impact on science, etc)

Goals/Motivation During the definition phases (0, A, B1) and the development of the mission, different configurations will be proposed going more and more deeply into the details of the intstrument and the pre-processing. For each configurations, a quick evaluation of the science performance will be needed. In order to do so, a pipeline or a set of pipelines should be developped as soon as possible.

- Quickly estimate the science performance of a given configuration of the instruement regarding each observation requirements defined in the Science Requirement Document
- Refine and update the precision in the estimation of the science performances, integrating in a compact way the output of various studies

Outputs

- Well defined figures of merit and a pipeline or a set of pipeline providing a "complete" estimation of the science performance of a given configuration in less that few days
- Document complementary to the SRD (so partially provided and validated by the SST) defining the quantities used for assessing science performance, the rationale behind them and current implementation

Time-frame

- First version required by the end phase 0
- Usable and complete version for early phase A (assessment of the input configuration to the ITT and of the first preliminary outputs)
- Evolution during phase A and B1 to integrate more precise methods and possibly new quantity defining for defining the science performances

Possible sub work-packages

- One workpackage per Science Objective or per source

Dependencies

- WP3.1, WP3.2 and WP3.3

2.8 Tools for generating and managing data quality indicators

Goals/Motivation

- It will be absolutely necessary to identify problems with the LISA data. There will be different levels of data quality problems. In the worse case scenario, there may be problems that are so severe that the data must not be used. On the other extreme, the highest quality data will still have occasional noise transients (glitches) or spectral lines. For the LISA signal searches to be successful data quality flags must be defined, as well as vetoes for specific time periods and frequency bands.

In addition to the LISA strain (h(t)) channels, it will be important to monitor as many auxiliary channels as possible. The placement of physical and environmental monitors (PEMs: accelerometers, magnetometers, thermometers, etc) will also be important to have. It will be possible to make statistical and modelled (where appropriate) associations between noise in the LISA h(t) channels and signals is the auxiliary channels and the PEMs.

Outputs

- Data quality flags defining the seriousness of the contamination of the data will be a necessary output. Various levels of data quality will be stated. Short duration times to be vetoed will also be provided. Spectral bands that contain excessive noise will be identified. The noise power spectral density will be modeled and estimated regularly, and compared to a regularly calculated noise budget.

Time-frame

- Many noise monitoring devices exist from LIGO-Virgo operations. For example, the means to identify noise transients, glitches, exist presently. Similarly, tools exist for identifying spectral noise lines. Tools also exist for making statistical associations between glitches that appear in an h(t) channel and an auxiliary channel. Similarly, software currently exists for calculating the coherence between an h(t) channel and an auxiliary channel.

  A comprehensive understanding of the ability to record and transmit data from auxiliary channels should be accomplished in the next two to three years. This same time period is probably also necessary for understanding what PEMs can be placed on the satellites.

  The methods for calculating the noise budget exist now, however the noise budget itself will evolve as the real experiment is constructed. A full understanding of the noise budget will likely not converge until the period of assembly, integration, verification and testing of the LISA systems is complete.

Possible sub work-packages

- Daily summary pages of various data quality indicators.

Dependencies

- The package will depend on the outputs from:
  Tools for detecting unmodeled signals
  Tools for modeling instrument performance
  Tools for assessing instrument/mission design choices

  The output of this package will flow to "Tools to assess impact of gaps, stationarity, data rates, spectral lines, glitches, etc on science output".

  The output of this package will flow to all of the subsection of "Low-latency pipelines".

2.9 Tools to assess impact of gaps, stationarity, data rates, spectral lines, glitches, etc on science output

Goals/Motivation Previous studies have assessed the impact of design trades such as arm-length, mission duration, orbit and noise levels on the science output. New studies are needed that go beyond the idealized approximation of stationary, Gaussian noise and consider the impact of gaps, non-stationarity, spectral lines and glitches. The new studies should also look at data rates and TDI variants. The studies will likely start with forecasting tools such as the Fisher Information Matrix, to be followed up with more detailed simulations.

- To quantify the impact of data gaps and realistic noise modeling on science output
- Come up with reasonable predictions for the number, duration and character of data gaps and disturbances
- Develop models for likely sources of non-stationarity and non-Gaussianity impacting the instrument noise

Outputs

- Tools to measure the impact on science, in particular parameter estimation and number of detections, of gaps, glitches, spectral lines *etc.*
- Strategies to mitigate the impact of data disturbances on the science output

Time-frame

- Our goal should be to have some preliminary studies completed during Phase A as the findings can impact the design
- Can build on heritage of earlier parameter estimation studies, but new elements are challenging to model. Likely will take several years.

Possible sub work-packages

- Develop a time-frequency domain likelihood function (essentially a noise model) that can handle gaps, non-stationarity and non-Gaussianity.
- Mission operations simulation to predict the frequency and duration of data gaps

Dependencies

- Waveform models (WP 1)
- Tools for assessing instrument/mission design choices (WP 2.7)
- Instrument response modelling (WP 3)

3 Instrument response modelling

The accessible science measurements of LISA are not directly h_+ and $h_\times$ of gravitational waves but the timeseries resulting of the pre-processing in particular of the application Time Delay Interferometry generators. If the noises are not considered, the three key ingredients are the instrument model itself, its orbits and the pre-processing with the application of TDI. Simulations will provide a numerical models of the response taking into account the complexity of the instrument. Simulations with noises and pre-processing reducing the noises will be necessary to provide a realistic instrument response. However simulations are not easy to integrate in search algortihms. Therefore an analytical approximations of the instrument response are needed, in time domain and frequency domain. To describe insrtument configurations, a clear parameterisation associated to the model should be provided. Table 3 summarises the work packages of this group.

WP	Description	Priority
3.1	Evolving instrument response model	0
3.2	Orbits and dynamics	0
3.3	TDI for DA (identify common aspects/inputs with TDI production)	0

Table 3: Work packages on instrument response modelling.

3.1 Evolving instrument response model

Goals/Motivation

- Having a numerical model (simulation) for computing signal in science measurements (phasemeters + necessary channels) corresponding to any GW described by a sky position and time evolving $h_+(t)$ and $h_\times(t)$ in Solar System Barysenter or by any diffuse GW sources (stochastic background).
- Having an analytical model for computing time evolution of the science measurements (phasemeter + other required channels) corresponding to any GW sources
- Having an analytical model for computing the science measurements in the frequency domains corresponding to any GW sources
- Having a time evolving PSD of noises in the phasemeter output (if not only the one in TDI is required)

Outputs

- Parametrisation of the instrument regarding the response to GWs
- Numerical model of the instrument response (simulators)
- Analytical models in time and frequency domain for computing the phasemeter measurements corresponding to any GW sources
- Global noises budget in science measurements

Time-frame

- A first instrument model is needed before the phase A to be provided with the ITT
- Parameters defining each new configuration appearing during definition phase

Possible sub work-packages No sub work-packages have yet been identified for this WP.

Dependencies

- WP3.2

3.2 Orbits and dynamics

Goals/Motivation

- Having numerical orbits for the reference configuration (the one provided as an input to the ITT) and possibly for others configurations obtained during the definition phases
- Having an analytic version of the orbits as close as possible of the numerical orbits
- Having a dynamical model of the 9 bodies (3 times 1 satallites and 2 test-masses)

Outputs

- Numerical orbits
- Analytic apporximation of the orbirts
- Model of the 9 bodies dynamics
- Simulations of the 9 bodies dynamics

Time-frame

- Numerical and analytical orbits with the model of the instrument before phase A
- 9 bodies dynamics during phase A

Possible sub work-packages

- Orbits
- Dynamics

Dependencies No dependencies yet identified.

3.3 TDI for Data Analysis (identify common aspects/inputs with TDI production)

It is necessary to pre-processed the telemetred data not only for calibration, clock resyncronisation, forces removal but also to reduce clock noises, spacecraft motion, laser noises. There are two steps in the application of TDI : the TDI steps 1 with a reduction of clock noises, spacecraft motion, half of laser noises and the TDI step 2 with the application of TDI generators to clean the rest of the laser noises.

Goals/Motivation

- Having an analytical formulation of the TDI steps 1 which are needed for reduction of clock noises, spacecraft motion and half of laser noises
- Having a common definition of TDI step 1 and generators (conventions)
- Having a self-standing pipeline to apply TDI given timeseries of phasemeter measurements and armlength
- Having an analytical model of full TDI chain including the TDI step 1 if it influence the response to GW
- Having a model for noise PSD in TDI channel: stationnary or time-evolving (simulation outputs)

Outputs

- Agreed convention for the TDI generator
- Self-standing pipeline to apply TDI
- Analytic model of TDI
- Model (numerical and/or analytical) for noise PSD in TDI channel: stationnary or time-evolving

Time-frame

- First numerical and analytical models ready with the model of the instrument before phase A
- First noise PSD for TDI outputs
- Evolving response of the instrument and the PSD of noises during the development of the mission with the evolution of configuration

Possible sub work-packages No sub work-packages have yet been identified for this WP.

Dependencies

- WP 3.1 and WP 3.2

4 Low-latency pipelines

Certain types of LISA source may be visible to electromagnetic telescopes, which means we need the capability to rapidly identify source candidates in the instrument data so that alerts can be sent out to EM partners for follow-up. In addition, there is good scientific motivation for trying to detect the gravitational wave emission generated during the merger phase of the system. A mechanism must be in place to trigger protected observing periods in advance, to ensure the existence of good quality data for the science analysis. Understanding the performance of the instrument in real time is also important, so that the general quality of data can be checked and any problems on the satellite identified quickly. The tools in this work package are designed to generate and distribute alerts by monitoring the LISA data in real time, as well as monitoring the instantaneous data quality from the instrument.

Table 4 summarises the work packages of this group.

WP	Description	Priority
4.1	Create low latency pipeline to run on ‘realistic’ data	3
4.2	Alert generation	3
4.3	Generation of data quality metrics and flags	2
4.4	Source-based observatory diagnostics	2
4.5	Search for unmodelled signals	2
4.6	Assessment and triggering of protected periods	1

Table 4: Work packages on low-latency pipelines.

4.1 Create low latency pipeline to run on ’realistic’ data

Goals/Motivation The LISA instrument will produce an evolving data set, and the data analysis will likewise need to produce an evolving fit for the many signals in the data. Low latency, “quick-look” analyses that fold in the latest data are vital for monitoring the instrument performance and generating alerts for possible electromagnetic follow-up of transient signals. The analysis will need to be able to handle gaps and data disturbances, and characterize the instrument noise.

- Noise characterization
- Low latency identification of the loudest signals
- Detection of short duration signals
- Localization of black hole mergers

Outputs

- Robust characterization of non-stationary, non-Gaussian noise in data with gaps and disturbances
- Rapid identification of short duration signals
- Initial solution for the global fit
- Refined localization of black hole mergers

Time-frame

- Much of this new territory. Explore multiple approaches over the next several years

Possible sub work-packages

- Methods to handle gaps, data disturbances and to characterize non-stationary and non-gaussian noise
- Methods to rapidly refine the sky localization of massive black hole binaries
- Methods to detect gravitational wave burst

Dependencies

- Waveform model (WP 1)
- Fast waveform generation (WP 1.7)
- Instrument performance monitors and data quality (WP 2.8, 2.10)
- Instrument response modeling (WP 3)

4.2 Alert generation

Goals/Motivation To maximize collaboration with the wider astronomy community new and updated LISA sources must be rapidly and predictably communicated with EM observers. For short-lived transients and in the late stages of SMBH mergers, low-latency source characterization and localization tools to get EM observers on source with minimal delay are of paramount importance.

Outputs

- Tools for rapid LISA source localization software for short-lived transients.
- Tools for low-latency updates to SMBH source localization in the late stages of the merger.

Time-frame

- Successful algorithms developed for LIGO/Virgo can be adapted to the LISA case. Prototypes need to be available for 'time domain' LDCs in $\sim 2-3$ yr.

Possible sub work-packages

- Tools for rapid LISA source localization software for short-lived transients.
- Tools for low-latency updates to SMBH source localization in the late stages of the merger.

Dependencies

- Significant overlap with Alert Distribution Tools WP 2.4.

4.3 Generation of data quality metrics and flags

Goals/Motivation

- While it would be helpful to have a distinction as clear as "the data is good" verses "the data is bad", there will always be situations where the quality of the data falls somewhere in between. Data quality flags will be used to provide an indication as to the quality of the data for specific periods of time, and to explain what are the problems. These flags should be produced in near real time.

  Through the design, construction, assembly, verification and testing a thorough understanding of the expected performance of the experiment should be achieved. During the operation of LISA it will likely be the case that the experimental performance and noise sources will change with time. As such, the noise budget must be updated frequently (probably daily). Known noise sources should be cataloged.

  The data quality metrics and flags will affect the signal searches. A good coupling must exist between the observed detector data quality and the signal searches. An understanding of the data quality will also be used to understand sources of noise, and to determine if there are methods to eliminate the noise sources.

  Characterization of the data will also allow for modeling of the noise. This could possibly provide the means for noise subtraction by various methods.

Outputs

- A catalog of noise sources.

  A record of the data quality as a function of time.

  Data quality flags describing the severity of noise problems.

  A daily summary page containing numerous and important data quality indicators.

  Cleaned data.

Time-frame

- Through recent work with LIGO and Virgo there are already many tools that exist for examining and evaluating data quality. Adaption of these tools to LISA applications can happen in the next year or two, hopefully on the same timescale as the development of the new Mock LISA Data Challenge.

  Methods for data cleaning exist now in the LIGO-Virgo community, and these could be adapted for LISA applications on the same timescale as the development of the new Mock LISA Data Challenge.

  Many data quality metrics will likely not converge until the completion of the assembly, integration, verification and testing phase of the satellites.

Possible sub work-packages

- Noise power spectral densities and glitch rates for the h(t) channels and for appropriate auxiliary channels.

Dependencies

- The package will depend on the outputs from:
  Tools for generating and managing data quality indicators
  Tools for detecting unmodeled signals
  Tools for modeling instrument performance
  Tools for assessing instrument/mission design choices

4.4 Source-based observatory diagnostics

Goals/Motivation General idea is to use verification binaries (VB) as a tool understand the instrument.

- Answer the question: Can verification binaries be used to improve upon the TDI ranging that uses pseudo random noise (PRN) modelling?
- How do the (expected) presence and absence of these binaries in various TDI observables complement the PRN-based TDI ranging? The residual signal in the null stream (GW free TDI observable) can be to estimate the ranging error. If the error in the ranging due to the SNRs of VBs are smaller compared to the PRN ranging, then we can optimise the PRN. As a first study to do is to look at what ranging errors can be derived from VBs alone. Then try to combine PRN and VBs.
- How can VBs be used to validate the calibration of the amplitude and phase of the signals from LISA? Given the known parameters from a VB parameters and the LISA orbit and it's geometry (and errors arising from those parameters) we have an expected strain with an associated ucertainty. Use this to compare the measured strain and uncertainty when we have the real data to compare against the expected/modelled quantities. The goal is to answer whether some of the strongest VBs can be used to calibrate the amplitude and phase of the signal. The same can be done with newly discovered galactic binaries those which are much brighter than the VBs. Their parameters could be measured relatively quickly and in turn this can be used for calibration validation.

Outputs

- Pipline that uses VBs as prior in an optimal estimation method to determine whether adding information from the VBs' signals changes TDI ranging.
- Pipeline for amplitude and phase calibration for a strain and uncertainty of the VB.

Time-frame

- 1-3 years The use of VB to in TDI ranging can be input for TDI algorithms
- 2-5 years amplitude and phase calibration tools

Possible sub work-packages No sub work-packages have yet been identified for this WP.

Dependencies

- WP1.4
- WP3.3

4.5 Assessment and triggering of protected periods

Goals/Motivation The goal here is to identify the mechanisms and decisions involved in triggering a protected period on the observatory. We also need to look at the constraints coming from the operations and the instrument itself.

- Detect SMBHBs days to weeks prior to merger and ensure the observatory remains in science mode (as much as feasibly possible) throughout the merger and ringdown.
- Routine interruptions need to be scheduled in such a way as to allow them to be rescheduled to avoid the protected period. This could imply advancing, or delaying of routine maintenance.
- Upon detection of a SMBHB merger, the decision chain needs to be identified (DPC informs SOC informs MOC, etc).
- Study currently known maintenance activities and look at scheduling those to allow for rescheduling.
- What is the impact on MOC/SOC operations?
- Who are the different players involved in the decision making process?
- What are the worst case scenarios (shortest notice)?
- Can we identify routine cases?
- How long does a protected period need to be? (how long is the ringdown etc)

Outputs

- An identified procedure against which protected periods can be triggered.
- Procedure to reschedule planned maintenance, and decision making chain to advance or delay maintenance.
- Identification of the tools needed to enable and communicate such events.

Time-frame

- This is a long term activity but needs to be integrated in the MOC/SOC development from the beginning.

Possible sub work-packages No sub work-packages have yet been identified for this WP.

Dependencies No dependencies yet identified.

5 Individual and global source identification codes

Science with LISA will rely on the identification of as many of the sources present in the data stream as possible, and on understanding any inter-dependencies in the inferred values of the parameters arising from confusion between the signals in the data set. As LISA sources are typically long-lived and present for a significant fraction, if not all, of the mission lifetime, a separate identification of individual sources or source types is not efficient. Instead, we will be attempting to carry out a simultaneous global fit for the parameters of all the sources in the data. This global fit will be continuously updated as additional data is received from the satellites. In practice, the global fit will be supported by additional codes which dig more deeply into the data and are focussed on particular source types. This codes will operate on the most recent best-fit residual from the global fit and will feed back any identified sources to be included in future refinements of the global solution. The WPs described in this section are designed to build this data analysis infrastructure. WP 5.1 is designed to build the algorithms and codes that will generate the global fit solution, while the remaining WPs will build the supporting codes identifying different source types, and evaluating instrument data quality.

Table 5 summarises the work packages of this group.

WP	Description	Priority
5.1	design of global fit strategies, including definition of output products	1
5.2	detection and parameter estimation of GBs	1
5.3	detection and parameter estimation of MBHBs	1
5.4	detection and parameter estimation of EMRIs	1
5.5	detection and parameterisation of unmodelled sources	2
5.6	detection and parameter estimation of IMRIs	3
5.7	detection and characterisation of stochastic backgrounds	1
5.8	production of cleaned (of sources) TDI variables	1
5.9	detection of modelled transient sources	2
5.10	identify data quality indicators and vetoes, and observatory diagnostics	2
5.11	develop instrumental 'noise' cleaning procedures	2

Table 5: Work packages on individual and global source identification codes.

5.1 Design of global fit strategies, including definition of output products

Goals/Motivation LISA data analysis presents the unique challenge of numerous overlapping signals. Hierarchical methods based on identification and subtraction of the brightest remaining sources accumulate errors which will bias or prohibit detection of weaker sources. Instead, the main production data analysis pipelines must perform a joint fit to all sources in the data and noise, an algorithmically and computationally challenging task. While different source classes will have dedicated pipelines, the global fit pipeline will provide the interface between the different source pipelines and orchestrate joint updates to all source and noise parameters to prevent the analyses from being harmed by avoidable biases and source confusion. The guiding goals are to:

- Develop a joint detection and characterisation pipeline for all sources present in the data.
- Define computationally efficient strategies for incorporating new data and updating the global fit.

- Standardize conventions for how to store information about detections needed to construct the source catalogs.

Outputs

- Top-level pipeline that provides the interfaces between individual source and noise analyses, and coordinates joint updates to all source/noise models.
- Standards for how to report source parameters with uncertainties, and detection confidence, for all astrophysical signals present in the data.

Time-frame

- The global fit pipeline will be developed in step with the LDCs.
- Prototype demonstrations of interfaces between pairs of source-classes in ∼3 yr.
- Prototype demonstrations of global analysis using intermediate versions of individual source pipelines, including production of mock catalogs, in ∼5 yr.
- Final version required L-2y.

Possible sub work-packages

- Prototype algorithms of pair-wise fits to identify incompatibility between single-source pipelines early in their development, and evaluate degree of covariance/confusion between different source types.
- Development of parameterized realistic noise model to incorporate into global fit pipeline.

Dependencies

- Data analysis tools WP 2.1-2.3
- Instrument response modelling (TDI) WP 3.3
- Individual source identification codes WP 5.2-5.7 and dependencies therein
- Source catalogs WP 6.1

5.2 Detection and parameter estimation of GBs

Goals/Motivation

- Develop one or more pipelines for joint identification and characterisation of all resolvable GB signals.
- Develop one or more pipelines for detailed analysis of single GBs (could be a use-case of the full search pipeline).
- Pipeline(s) must interface with global fit pipeline, operate on raw data or latest residuals from the global fit, and return a list of the GB's source parameters, uncertainties, and detection confidence.

Outputs

- Pipeline for GB detection and characterization of full galaxy which will interface with global fit pipeline.
- Pipeline for detailed study of isolated GBs.

Time-frame

- Pipeline(s) will evolve with LDCs, prototype algorithms demonstrated within a year.
- Mature algorithms adapted to realistic data scenarios in ∼5 yr.
- Final version required L-2y.

Possible sub work-packages

- GB contribution to global fit pipeline
- Dedicated single-source GB pipeline for detailed analysis (including all relativistic and astrophysical effects).

Dependencies

- Waveform modeling WP 1.4 and fast waveform generation WP 1.8
- Data analysis tools WP 2.1-2.3
- Instrument response modelling (TDI) WP 3.3

5.3 Detection and parameter estimation of MBHBs

Goals/Motivation

- GIven the degeneracies on the parameters observed in old-days MLDC we need to develop more than one pipeline for identification and characterization of MBHB signals
- The pipelines should be split and/or coupled with GB detection/characterization
- The pipeline should be a part of the global fit either right away or be inserted into the global fit with preliminary estimation of source parameters with the low latency.
- The pipeline should be flexible enough to utilize different/evolving models for MBHB signal and be robust to possible systematics in the signal modelling (marginalization over uncertainties).
- Pipeline delivers catalogue of source (both over and sub-threshold) with associated uncertainties (both in numbers and parameters)

Outputs

- Several stand-alone pipelines running in parallel for cross-checking, or a single pipeline utilizing several methods.
- Pipeline incorporated in the global fit analysis, interacting with other pipelines (or being a part of one global pipeline).

Time-frame

- Pipeline will be developed within LDC, the first version should be available within a year.
- Robust pipeline which is able to recover datectable (loud enough) MBHB in presence of other major sources and non-stationary noise $\sim$ 5y
- Final version L-2y

Possible sub work-packages

- Stand-alone pipeline(s) for detecting MBHB (see above)
- MBHB detection as apart of a global fit with/without input form the stand-alone pipeline

Dependencies

- Overlap with WP4.1, WP 5.1
- Depends on WP1.2, WP 1.8, WP2.1-2.3, WP4.3,4.4, WP6.1, 6.2

5.4 Detection and parameter estimation of EMRIs

Goals/Motivation

- Develop one or more pipelines for identification and characterisation of EMRI signals.
- Pipeline should interface with global fit pipeline. It will operate on the latest residuals from the global fit and return a list of identified EMRI sources for subsequent incorporation in and refinement by the global fit.
- Pipeline should be able to use the best available EMRI model at the current time, but also be robust to waveform modelling uncertainties, e.g., by including marginalisation over waveform model uncertainties.
- Pipeline outputs are catalogues of source parameters and uncertainties.

Outputs

- A pipeline for EMRI identification and characterisation.

Time-frame

- Preliminary pipelines for EMRI characterisation were developed during the MLDCs. Existing pipelines have not been shown to be robust to modelling uncertainties or source confusion.
- Improved and robust pipelines need to be developed on a $\sim$ 5y timescale and verified within the MDCs.
- Final versions required L-2y.

Possible sub work-packages Sub work-packages would be built around different approaches to EMRI data analysis.

- Stack-slide EMRI search pipeline. Exhaustive EMRI search using templated search for short ($\sim$ 2 wk) sections of EMRI signals, later combined incoherently.
- Stochastic EMRI search pipeline. EMRI search based on tuned MCMC algorithm, as used in the MLDCs.
- Semi-coherent EMRI search. First identify individual EMRI waveform harmonics using phenomenological templates, then combine to identify sources and increase significance.

Dependencies

- EMRI waveform modelling (WP 1.2). Search will use best available templates, or approximations tuned to match these.
- Common data analysis framework and tools (WP 2.1, WP 2.2, WP 2.3).
- Fast waveform generation tools (WP 1.7). The EMRI search will rely on fast but accurate approximations to EMRI waveforms.
- Instrument response modelling (WP 3.1) and data quality indicators (WP 4.3, 4.4).
- Global fit strategy (WP 5.1). EMRI search will operate on cleaned data provided by global fit pipeline and return EMRI catalogues for incorporation in global fit.
- Catalogue structure and interfaces (WP 6.1, 6.2). The output from the EMRI search pipeline must include the data products required for cataloguing.

5.5 Detection and parameterisation of unmodelled sources

Goals/Motivation The 2010 *New Worlds New Horizons* Decadal survey remarked that "It would be unprecedented in the history of astronomy if the gravitational radiation window does not reveal new, enigmatic sources." To be in a position to make such discoveries we need techniques that can detected unanticipated and unmodelled signals. The task is complicated by the fact that LISA is one instrument, so the coherent network analysis techniques used to search for unmodelled signals in the LIGO/Virgo data are not immediately applicable. The signal-insensitive Sagnac channel, and differences in the transfer functions for signals and noise will key to any detection strategy.

- Develop techniques that can distinguish unmodelled signals from instrumental artifacts
- Develop techniques to characterize unmodeled signals

Outputs

- One for more analysis tools for extracting unmodelled signals
- Tools to characterize unmodelled signals (frequency content, time evolution, duration *etc*)

Time-frame

- This is largely uncharted territory. Good to explore multiple approaches over the next several years
- Short duration transients are likely the easiest target. May be able to repurpose LIGO/Virgo burst algorithms in near term

Possible sub work-packages

- Transfer functions for all conceivable noise transients
- Generic waveform reconstruction techniques, possibly wavelet based

Dependencies

- LISA Instrument Simulator (WP 3.1)
- TDI models for noise transients (WP 3.3)
- TDI models for generic signals (WP 3.3)

5.6 Detection and parameter estimation of IMRIs

Goals/Motivation

- A black hole of intermediate mass ($M \sim 10^2$–$10^4 M_\odot$) (IMBH) inspiraling into a massive black hole will appear as an IMRI in LISA. The existence of IMBHs is uncertain, but detection of one or more IMBHs would be highly significant.
- We separate IMRIs from EMRIs because waveform models are more uncertain for mass ratios in the IMRI range, $\sim 10^{-4}$–10^{-2}. Handling waveform uncertainties will be crucial in this search.
- This pipeline will interface with the global fit pipeline in a similar way to the EMRI pipeline (WP 5.4).

Outputs

- A pipeline for detection and characterisation of IMRIs.

Time-frame

- Due to the speculative nature of these sources, this is not a priority work package. After development of searches for MBHBs (WP 5.3) and EMRIs (WP 5.4), the efficiency of those searches for detecting IMRIs should be assessed.
- Further tuning and optimisation of an IMRI search will continue during mission development and tested within the LDCs.
- Final version required L-2y.

Possible sub work-packages Sub work-packages would be built around different pipelines, as in the EMRI case, e.g.,

- Stack-slide IMRI search.
- Stochastic IMRI search.
- Semi-coherent/time-frequency IMRI search.

Dependencies

- IMRI waveform modelling (WP 1.5). Search will use best available templates, or approximations tuned to match these.
- Common data analysis framework and tools (WP 2.1, WP 2.2, WP 2.3).
- Fast waveform generation tools (WP 1.8).
- Instrument response modelling (WP 3.1) and data quality indicators (WP 4.3, 4.4).
- Global fit strategy (WP 5.1). IMRI search will operate on cleaned data provided by global fit pipeline and return IMRI catalogues for incorporation in global fit.
- MBHB and EMRI searches (WP 5.3, 5.4). IMRI search will most likely be derived from searches for other source types.
- Catalogue structure and interfaces (WP 6.1, 6.2). The output from the IMRI search pipeline must include the data products required for cataloguing.

5.7 Detection and characterisation of stochastic backgrounds

LISA data are expected to contain several contributions from stochastic foregrounds/backgrounds – of astrophysical and/or cosmological origin – including (i) the foreground from WD-WD binaries at mHz frequencies, (ii) a possible foreground from EMRIs in the same frequency range, (iii) a foreground from stellar-mass black holes at $\sim$ 0.01 Hz, (iv) a foreground from neutron stars in the frequency range of the stellar-mass black holes, and (v) a background from processes in the early Universe whose level is unknown.

Goals/Motivation

- To detect and characterise stochastic signals present in LISA data.
- To characterise the properties of the stochastic signals present in the data (the signals may not be Gaussian and/or stationary and/or isotropic)

Outputs

- Pipeline(s) for the detection and characterisation of stochastic gravitational-wave signal(s).
- Given the nature of the problem, and the lack of an existing end-to-end approach that deals with conditions that are remotely realistic, it is highly desirable that multiple approaches and pipelines are developed early on to explore performances and trade-offs.
- The pipeline ability to identify and characterise the signals will depend on the TDI outputs and links available in the configuration, and will need to work under the different conditions (level of stationarity/Gaussianity of the noise, data drop-outs and corruption) and configurations. Depending on which links are available, the ability to recover stochastic background could be totally compromise. An important output of this WP in the early phase of development is to identify the minimal instrumental conditions/configuration for the recovery of a stochastic signal.

Time-frame

- A preliminary pipeline was developed during the MLDCs. However the operation conditions were highly idealised (Gaussian and stationary instrumental noise, primarily full 6 links available in the LISA instrument).
- Final version required L-2yr.

Possible sub work-packages

- Gaussian/stationary background signal vs more generic backgrounds
- Isotropic background
- Anisotropic background
- Physical characterisation of multi-component background/foreground based on de-fault (TBD) models
- A possible additional hierarchy of the WPs is to have workpackages for the "full-configuration" (6 links) and as for each of them sub-workpackages that deal with "reduced-configurations" (less than 6 links)

Dependencies

- This workpackage depends on pretty much everything else, as it deals with the analysis of "what is left" after all the individual sources have been resolved. (add numbers)
- The workpackage depends on all the data-characterisation workpackages (add numbers)
- It depends on all the WPs that deal with identification of resolvable sources, global fit, cleaned data, TDI reconstruction (I will need to add numbers later)

5.8 Production of cleaned (of sources) TDI variables

Goals/Motivation While the global fit is a necessity for the final catalog of sources, detailed analyses of individual sources will not want to be burdened by the full set of global-fit parameters. Examples where the analysis would want to focus on a single source include investigations of relativistic and astrophysical effects beyond the standard waveform models, or low-latency updates of merging SMBH parameters to be communicated with EM observing partners.

- Provide robust global-fit subtraction tools including error budgets from source parameter uncertainties.
- Provide up-to-date cleaned TDI variables which can quickly be modified to focus on single sources for low-latency parameter updates.
- Provide clean residuals for instrument and noise characterization work.

Outputs

- Tools to produce residuals from the global fit with uncertainties.
- Tools to add back individual sources of interest for detailed/updated analysis.
- Error-budget from waveform subtraction for the cleaned TDI variables.

Time-frame

- These are straightforward tools they depend on waveform, search pipeline, and catalog development. Can be demonstrated in early LDCs.
- Final version L-2yr.

Possible sub work-packages

- Develop algorithms for incorporating uncertainties from signal subtraction into analyses of residual data.
- Develop tools to interface with catalog and add back sources to TDI variables for further analysis/updates to parameters.

Dependencies

- Waveform models WP 1.1-1.8
- Data analysis tools WP 2.1-2.3
- Instrument response modelling WP 3.1
- Individual and global source identification codes WP 5.1-5.6
- Source catalogues WP 6.1,6.2,6.5

5.9 Detection of modelled transient sources

Goals/Motivation

- Develop a pipeline for identifying short-duration modelled transient sources in the LISA data stream.
- Known potential transients are cosmic strong cusps and extreme-mass-ratio bursts.
- Pipeline must be as robust as possible to instrumental artefacts and hardware failures.

Outputs

- A pipeline for identifying, characterising and cataloguing modelled transient sources in the LISA data.

Time-frame

- This is a lower priority than other modelled source types due to the uncertain event rate.
- This search will be more sensitive to instrumental artefacts. Assessing robustness will rely on the development of the tools for characterising and simulating instrumental noise, so this must happen first.
- Preliminary pipelines needed on a $\sim$ 5yr timescale, to be verified in LDCs. Searches for cosmic strings were demonstrated in the MLDCs, but without contamination from instrumental artefacts.
- Final pipelines required L-2y.

Possible sub work-packages

- Pipeline for identifying extreme-mass-ratio burst signals.
- Pipeline for identifying cosmological burst sources. NB this is likely not to be a separate work package, but just the previous pipeline with a different waveform model.

Dependencies

- Waveform modelling (WP 1.7).
- Common data analysis framework and tools (WP 2.1, WP 2.2, WP 2.3).
- Instrument response modelling (WP 3.1) and data quality indicators (WP 4.3, 4.4).
- Search for unmodelled transients (WP 4.5). The search for unmodelled transients will be sensitive to modelled transients so the development of these two packages is linked. The pipelines could run completely independently or the unmodelled transient search could trigger the modelled search, with the modelled search also being used to find quieter signals.
- Global fit strategy (WP 5.1).This search will most likely operate on cleaned data provided by global fit pipeline and return lists of identified transients for incorporation in global fit. It could also be part of the global fit pipeline.
- Catalogue structure and interfaces (WP 6.1, 6.2). The output from the search pipeline must include the data products required for cataloguing.

5.10 Identify data quality indicators and vetoes, and observatory diagnostics

Observatory diagnostic allows to identify segments of bad data that can negatively affect the detection and analysis of GW signals. When segments of poor data quality are identified a data quality warning should be released, eventually veto flags should be issued in order to exclude the given segments from the analysis of GW signals. Data quality warnings and vetoes should be issued if the origin of the data quality degradation can be clearly identified.

Goals/Motivation

- Spurious signals (e.g. glitches, micro meteoroids impacts) can be mistaken for transient GW signals.
- Total or partial loss of data can occur because of downlink failure, sub-system failure, SC to SC link failure.
- Anomalies or malfunctioning in a subsystem can generate extra noise at the output of the instrument.
- Any anomaly in the data can affect the ability of reconstructing TDI variables.
- Data anomalies could impair the ability of extracting GW signals.

Outputs

- Identification of the indicators of the healthy status of the instrument.
- Identification of possible sources of instrument anomalies affecting the quality of the data.
- A set of pipelines monitoring the healthy status of the instrument.
- A set of pipelines generating data quality indicators.
- A set of pipelines generating veto flags.

Time-frame

- Investigation of data quality indicators should start within 1 year time frame.
- Pipelines development should start within 1 year. Pipelines should be tested in the framework of the LDCs.
- Fully functional preliminary pipelines are needed in 5 years.
- Final version of the pipelines are required at L-2y.

Possible sub work-packages

- Monitoring the healthy status of the optical metrology system. Analysis of the impact on science, data quality indicators and veto flags generation.
- Monitoring the healthy status of the GRS. Analysis of the impact on science, data quality indicators and veto flags generation.
- Analysis of the environmental monitors (radiation monitors, magnetometers, thermometers), impact on science, data quality indicators and veto flags generation.
- Detection of spurious signals (e.g. glitches, micro meteoroids), cross-correlation analysis of all available channels and their signature in TDI.
- Impact of instrument maintenance (e.g. TM discharge, telescope repointing, mode change) and instrument calibration (e.g. signal injections) on data quality.
- Vetoes generation as a consequence of instrument anomalies (e.g. total or partial loss of data, sub-system fault).

Dependencies

- Common data analysis framework and tools (WP 2.2, WP 2.3, WP 2.9).
- Instrument response modelling (WP 3.1)
- Generation of data quality metrics and flags (WP 4.3)

5.11 Develop instrumental 'noise' cleaning procedures

Goals/Motivation

- Achieving the optimal observatory performance will likely involve the removal of noise through post-processing.
- This will first rely on the identification of noise couplings from some auxiliary channels to the TDI outputs.
- for each identified noise contribution
  - the transfer function needs to be parameterised and identified, studied and monitored.
  - the source noise measurement needs to be monitored and maintained as part of the routine data set.
- for any such coupling mechanism, 'safety' to GW signal needs to be assessed. This could be achieved by searching for the presense of high-SNR signals in those channels, or by performing dedicated hardware injections. Look at what has been done in LIGO in this regard.

- for such couplings, the correct point in the signal processing chain to do the subtraction needs to be identified. For example, it could be done before TDI, or as part of the TDI application, or post TDI.
- some of these couplings will require dedicated experiments to measure appropriate parameters. Examples of this are calibration of the coupling of laser power fluctuations as a force noise on the test masses. Another would be the subtraction of SC jitter through tilt-to-length couplings.

Outputs

- List of expected couplings and noise subtractions based on the experience from LPF.
- List of 'new' couplings which were not studied in LPF.
- Designs of required calibration experiments to be performed in operations.
- Identification of analysis tools.

Time-frame

- This work should be looked at reasonably early, during Phase A, from the point of view of ensuring that, for the 'known' couplings:
  - the required auxilliary signals are made available
  - the required experiments can be performed from a hardware point of view
- in the longer term, simulations and hardware development may reveal new couplings which will need to be studied.

Possible sub work-packages

- Correlations and couplings of TTL through TDI (part of Simulation WG work)
- Correlations and coupling of RIN through TDI (part of Simulation WG work)
- Adaptation of LPF calibration experiments to LISA situation

Dependencies

- there is a link with the data analysis tools
- there is a strong link with the work plan of the Simulation WG

6 Source catalogues

The LISA catalogue is the top-level science product of the mission, together with high-profile publications and data releases accompanying discoveries and surveys. Even those will be supported by the catalogue, either through specialized products, or targeted "views." Thus, the catalogue will be the primary interface between the LISA mission (and by extension science team and the core user group contributing to searches) and the broader astronomical community, which will use the catalogue to derive and test, e.g., MBH and Galactic-binary population models. The catalogue will also support EM counterpart alerts.

Because of the global-fit nature of the LISA science solution, the catalogue will also be a crucial object internally within the LISA data system and science analysis. The internally facing catalogue is updated continuously by data-analysis pipelines to report new detections and revise parameter estimates, and it is accessed by the pipelines to perform partial waveform subtractions from the dataset.

Table 6 summarises the work packages of this group.

WP	Description	Priority
6.1	Definition and design of catalogue(s) and interface with global fit solutions	1
6.2	Tools for interfacing, visualisation, storage, searching, etc	1
6.3	Protocols for catalogue evolution, QA, and change tracking	3
6.4	Management of catalogue release, dissemination, etc	3
6.5	Cataloging of unmodelled sources	2

Table 6: Work packages on source catalogues

6.1 Definition and design of catalogue(s) and interface with global fit solutions

Goals/Motivation

- Design LISA catalogue as a database that supports several modes of use and different queries, ranging from the internal global fit to the needs of astronomers.
- Develop tools and methods to incorporate as much information as possible about source-parameter uncertainties (e.g., as Monte Carlo sample files) and about correlations between sources.
- Allow catalogue to be accessed efficiently for the purpose of reconstructing the data likelihood (or a partially marginalized/extremized likelihood) as a function of source parameters.
- Allow catalogue to be accessed efficiently for the purpose of building partially regressed data products (e.g., dataset minus best-fit detected SMBHs and loud Galactic binaries).
- Design catalogue in tandem with the LISA data-archive interface.
- Catalogue (or a catalogue extension) may allow also ingestion and access of data about EM counterpart searches and observations.

Outputs

- Catalogue design whitepaper.
- Catalogue specification document.
- Catalogue implementation as database (several versions support increasing functionality).
- Libraries and code examples to access and update catalogue (see also 6.3).
- Simulated source lists and records to populate prototypes.

Time-frame

- Simple prototype by third (?) LDC.
- Initial design by beginning of Phase A.
- Full design and prototype by end of Phase A.
- Mature design by end of Phase B.

Possible sub work-packages

- Single-source catalog entries.
- Multi-source interface and management of correlations.
- Multimessenger catalog.

Dependencies

- Catalogue will be updated or queried by many pipelines and tools across all WPs.
- Catalogue will depend on source description in file formats (WP 2.3), and may be partially implemented through those.

6.2 Tools for interfacing, visualisation, storage, searching, etc.

Goals/Motivation

- This WP amounts to implementing the basic functionality of the LISA data archive for internal and external query.
- The WP will provide software interface specification and implementation for pipelines to obtain/update entries from/to catalogue.
- The WP will provide tools, examples, and tutorials for GW and EM astronomers to query catalog, produce meaningful visualization of entries under various aggregations, etc.

Outputs

- Use-case/design document.
- Web/database service providing access to catalogue database (and multiple versions thereof).
- Web interface for human access, allowing complex/graphical queries.
- Visualization plugins for web interface and/or external applications.
- Scripting tools for automated access by pipelines.
- Simulated catalogues for testing.

Time-frame Follows 6.1 since it implements it operationally.

- Simple prototype by third (?) LDC.
- Initial design by beginning of Phase A.
- Full design and prototype by end of Phase A.
- Mature design by end of Phase B.

Possible sub work-packages

- Server.
- Web interface.
- Visualization tools.
- Scripting access tools.

Dependencies

- Catalog design 6.1.
- Common software tools 2.2.
- Data exchange formats 2.3.

6.3 Protocols for catalogue evolution, QA, and change tracking

Goals/Motivation

- New source entries will appear and will be updated as the dataset is collected, and as increasingly sophisticated analysis pipelines are applied to sources. The catalogue should fully support these insertions and updates.
- The catalogue should be capable of maintaining concurrent (and possibly discrepant) accounts of the same source, or even different number of sources.
- Accordingly, we should define an index of "catalogue record quality" and maintain fully accessible provenance information in order to support catalogue queries.
- Thus, the LISA catalogue should implement a sophisticated form of journaling and version control.

Outputs

- Use-case/design document.
- Implementation in database schemas and housekeeping code.
- "diff" and "log" tools (text-based and graphical) for catalogue entries and queries.
- Testing suite for QA.
- Simulated catalogues and catalogue updates for testing.

Time-frame

- Mature design by end of Phase B.

Possible sub work-packages See Outputs.

Dependencies

- WP 6.1 and 6.2

6.4 Management of catalogue release, dissemination, etc.

Goals/Motivation

- As much as the catalogue will be a living object, continuously updated throughout the mission and beyond, the needs of coherence and reproducibility suggest that the outward-looking catalogue should be made available to the astronomical community in regular incremental releases.
- Catalogue releases should always accompany new data products (e.g., TDI files), and will need to be properly "synchronized."
- We need to develop a plan for the releases, looking at examples from other missions and observatories, and gathering input from the astronomical and GW community.
- We need to develop requirements and quality controls for what is included in a release, as well as timelines and responsibilities to produce and approve them. Again, look for high-profile examples (e.g., SDSS).

Outputs

- Use-case/design document.
- Plans for accompanying publications, talks, etc.
- Web interface.
- Simulated releases (including very complete ones for "final" mock data challenges).

Time-frame

- Mid phase B.

Possible sub work-packages See outputs.

Dependencies

- WPs 6.1, 6.2, and 6.3.

6.5 Cataloguing of unmodelled sources

Goals/Motivation

- GW sources detected without direct reference to a physical waveform model require appropriate representation in the catalog, which will be married closely to the analysis that produced the detection: for instance, the catalog entry may represent a posterior distribution of wavelet coefficients.
- The catalog may consist directly of a posterior distribution of reconstructed waveforms.

Outputs

- Schemas and specification document.
- Software tools.
- Simulated example entries.

Time-frame

- End of phase A.

Possible sub work-packages See outputs.

Dependencies

- WPs 6.1, 6.2, 6.3, and (very closely) 8.12.

7 Multi-messenger, multi-band

To maximise the scientific return of the LISA mission, the synergy with other messengers (like light and particles) is of paramount importance. Likewise, we recognise the high gain of a synergistic use of low and high frequency GW detectors. The overall goal of this section is to create tools and infrastructures that, in synergy with WPs in session 8, will allow us to realise the full scientific potential of multiband/multi-messenger data. Particular attention is given to tool development and information sharing. Since our goals naturally entail reaching out to other scientific communities, outside the LISA consortium, this section presents WPs dedicated to communication strategy and policy (WPs 7.5, 7.6, and 7.7). The rest of the WPs deals with tool development and catalogue building, essential both before and after the LISA launch (WPs 7.1, 7.2, 7.3, 7.4): prior to real data acquisition, we will create mock data/catalogues to guide the development and to test analysis pipelines and information sharing infrastructures; on the other hand the work to be performed by these WPs will be essential to actually calibrate the instrument and exploit the incoming data after the LISA launch.

Table 7 summarises the work packages of this group.

WP	Description	Priority
7.1	Exploration of multi-messenger science with LISA	1
7.2	Joint analysis methods/tools for EM and GW	2 or 3?
7.3	Identification of VBs from EM catalogues	2 or 3?
7.4	Multi-band GW analysis	1 or 2?
7.5	Establish partnerships and MOUs etc	1
7.6	Define data/communication protocols	2
7.7	Technical outreach and communication	2

Table 7: Work packages on multi-messenger and multi-band astronomy.

7.1 Exploration of multi-messenger science with LISA

This WP team will perform an *exploratory* investigation of multi-messenger science for all LISA sources except for VBs, which have a dedicated WP (WP7.6). The investigation is concerned with both LISA discovered sources (that we will call "LISA triggers") and LISA source candidates discovered in E.M. or by particle detectors ("LISA follow-up sources") that can be potentially discovered in the LISA data stream below the detection threshold. The work will be performed during phase A and will set the stage for the development of numerical tools and a joint analysis pipeline described in WP7.5.

Goals/motivation

The WP is focused on the *assessment* of the work to be performed in order to identify:

- the most promising E.M./particle counterparts to LISA astrophysical sources, given the available instruments simultaneously online with LISA;
- the best strategy to obtain and exploit redshift information from GW sources in order to measure cosmological parameters;
- possible E.M./particle counterparts to the stochastic background (e.g. predicted contribution to X-ray or Infrared background): this item is stated separately because both the GW and the E.M. overall investigation (observation and modelling) is different from that of individually detected sources;

- inform the design of a LISA data analysis and parameter estimation mode dedicated to *prompt* E.M./particle counterpart identification.

In particular, we need to quantitatively address the points described below in “Outputs”.

Outputs

We will produce a document stating:

- the synergistic measurements needed to maximise the science return;
- the amount of work required;
- the expertise needed;
- team membership: identify inside and (currently) outside the LISA consortium researchers who can valuably contribute;
- an overall work timeline,

to successfully reach our aforementioned goals ahead of the LISA launch.

Time-frame
We envisage three phases:

- first phase, month 1 and 10: i) produce an initial list of all the proposed counterparts and a list of all the facilities, available before and during LISA operation, that can potentially be exploited; ii) analytically assess the feasibility to observe these E.M. counterparts given the instruments available; iii) set priorities for second phase.
- second phase, month 11 and 21: design and perform basic simulations of E.M. counterparts with or without a LISA trigger.
- third phase, month 22 and the 24 month: writing final report.

Sub work-packages

The work will be divided among subgroups dedicated to a particular GW astrophysical source + an exploratory sup-WP. There can be an overlap between members of different sub-WPs. Here are the proposed sub-WPs:

- SMBHB + SOBH
- EMRI
- Stochastic background
- Unknown counterparts: this sub-WP will explore the best observational strategy, regardless of the theoretically proposed counterparts.

Dependencies

Mutual dependency is envisaged between WPs dealing with multi-messenger/multi-GW band signals, in addition to those related to alert distribution tools and low-latency. Specifically:

- all the WPs in Section 7
- WP described in Subsection 2.4
- all WPs in Section 4
- science exploitation (WPs in session 8, expect 8.1, 8.2, 8.7)

7.2 Joint analysis methods/tools for EM and GW

This WP mainly capitalises on the work done previously in WP7.1. Its main objective is to creates the numerical tools to exploit the full scientific potential of the GW events by adding the EM information. These tools should be built so that can be potentially used by all consortium members.

Goals/Motivation

- develop codes to simulate the main physical observables (e.g lightcurves, spectra) for selected counterparts to GW sources, following the preselection done by WP7.1. The outputs do not take into account instrument response, source distance and extinction from intervening matter. Those will be accounted for in the next item;
- develop an interface that returns observables (i.e. fluxes, photometric information, observed spectra, observed redshift, etc...) when given as an input the source physical parameters and the specification of an instrument/observational strategy;
- develop a pipeline to analyse GW data including information from counterparts, in a fully consistent statistical framework;
- inform the design of a LISA data analysis and parameter estimation mode dedicated to *prompt* E.M./particle counterpart identification.

Outputs

- a modular simulation code for each selected E.M. counterpart: simulation code shall be modular, so that new counterparts, observables and instrument specifications can be easily added (e.g. as plugs-in);
- an interface to exploit these codes aforementioned and return observables;
- a joint analysis pipeline;
- documentation for all developed software;
- a document with all suggested specifications for a prompt LISA data analysis and parameter estimation mode.

Time-frame We envisage a 3-phase time-plan:

- first phase, month 0-24: construction of counterpart simulation code with user interface and building of a mock data catalogues;
- second phase, month 12-36: assemble and test a joint GW/EM analysis pipeline, exploiting mock data; outline specification for a prompt LISA data analysis and parameter estimation mode;
- third phase, month 37-42: overall document writing.

Possible sub work-packages

- Counterpart simulation team.
- Joint GW/EM analysis team.

Dependencies

- low latency pipeline and alert generation (WP 4.1, WP 4.2);
- multimessenger science definition (WP 7.1);
- MOU definition and data sharing protocols (WP 7.5, WP 7.6);
- science exploitation (WPs in session 8, expect 8.1, 8.2, 8.7)

7.3 Identification of VBs from EM catalogues

Goals/Motivation
Tight Galactic white dwarfs binaries can be guaranteed GW sources for LISA. This population of persistent GW sources can be observationally characterised prior to the LISA launch. This data set is important for both instrument calibration and science exploitation. The main goals of the WP are the following:

- identify the specific calibration goals in collaboration with WP 4.4;
- identify the specific science goals in collaboration with WP 8.4;
- identify the minimum number of VBs and the threshold measurement accuracy needed to reach our science and calibration goals;
- identify the best observing campaigns to acquire the above sample before LISA flies;
- ensure that the above sample is gathered and fully analysed: this can be partially done by the WP team and/or in collaboration with other groups.

Outputs Our mail outputs will be:

- a document reporting detailed specification of number of VBs and precision measurement needed in order to achieve the calibration and science goals ;
- several mock catalogues (for e.g. GAIA, LSST) which allow us to determine the minimum requirements to fulfil our goals and to plan observing campaign (goals 3 and 4 above); These can/should also be used to test analysis pipelines (for both calibration and science outputs);
- a well characterised E.M. catalogue of VBs and their *expected* GW properties;

Time-frame The time-frame (to be refined when the priority of this WP is settled) is

- first phase: identify science and calibration goals and build mock E.M. catalogues to identify the best observing strategy and a data acquisition plan;
- second phase: a first version of the (real) E.M. and GW catalogue for VBs should be delivered;
- third phase: finalise catalogues.

Possible sub work-packages

- sub-WP in charge of mock data;
- sub-WP in charge of real data acquisition and analysis;
- sub-WP in charge of GW characterisation of VBs to be added to the E.M. catalogue;

Dependencies

- WP 4.4 (calibration tools)
- WP 8.1, 8.2 (science exploitation tools)

7.4 Multi-band GW analysis

LIGO-Virgo detections of massive SOBH provided a new class of objects observable in different GW bands, from mHz (LISA) to kHz (LIGO-Virgo). LISA will observe several SOBH in the relatively early inspiral phase, providing complementary information to ground based observations, for example by measuring eccentricity and possibly environmental influences and spins. Moreover, most SGWBs are predicted to span several orders of magnitude in frequency. Possible multi-band detections of a SGWB by LISA, LIGO-Virgo and PTA are also crucial for disentangling the origin of the signal (astrophysical and/or cosmological) and for imposing constraints on early universe and fundamental physics scenarios (such as the presence of phase transitions leading to cosmic strings, the inflationary model and so on.)

Goals/Motivation

- send alerts to ground based detectors;
- use, when possible, ground based observations for targeted searches of sub-threshold events in the LISA data stream;
- exploit at best the potential of joint space- and ground-based observations:
  1. refine source parameters,
  2. test GR and alternative theories of gravity,
  3. combine information from both bands to enhance astrophysical inference;
  4. test the presence of correlated SGWBs in several bands and infer constraints on their origin.
- develop tools for joint data analysis and joint inference from ground and space based information.

Outputs

- mock data catalogue of multi-frequency sources;
- data analysis pipelines for joint LISA and ground based detections:
  1. refinement of source parameters,
  2. determination of possible deviations from GR (e.g. dipolar radiation).
- inference pipelines from LISA data and from LISA/ground based data:
  1. astrophysical inference on the SOBH population (i.e. disentangle different formation channels),
  2. tests of GR and alternative theories of gravity,
  3. use of SOBHB as standard sirens for cosmology,
  4. analyse possible correlated SGWB signals.
- pipeline for targeting LISA sub-threshold events, following LIGO triggers.
- publication of several technical papers describing pipelines and tests on mock data (e.g. mock studies on enhancement of parameter estimations, tests of GR, astrophysical inference).

Time-frame

- month 1-24: First version of single source DA pipelines for testing on LISA mock data.
- month 12-36: First version of joint inference (astrophysics and tests of gravity).
- final DA and joint inference pipelines required by L-y1.

Possible sub work-packages

- Sub WP 2: Development of joint DA pipelines.
- Sub WP 3: Development of joint inference pipelines.

Dependencies

- SOBHB waveforms (WP 1.6).
- Framework for common data analysis tools (WP 2.1, WP 2.2, WP 2.3).
- Individual and global source identification codes (WP 5).
- Partnership and communication protocols (WP 7.2, WP 7.3, WP 7.4).
- Source catalogues (WP 6).
- Key scientific projects: cosmological parameters (8.10), tests of GR (8.7), SOBH populations (WP 8.9).
- Collaboration with population synthesis modellers, especially for astrophysical inference from data.
- Collaboration with the LVC.

7.5 Establish partnerships and MOUs etc

Goals/Motivation

E.M. counterparts to GW detections can arise on a quite broad range of timescales and wavelengths. In particular, to be able to detect E.M. signals which are broadly speaking coincident (from seconds to a few weeks or even precursors) with the GW signal requires a prompt alert from the LISA consortium, *well before* any official publication of the GW event is possible. This poses a challenge: on the one hand the LISA consortium wishes to maximally facilitate the discovery of counterparts to our GW sources; on the other we would also like to get full credit for our discoveries while being able to publish only when full confidence in our data analysis results has been reached. The same is true for GW mutli-frequency observations.

It is therefore necessary:

- to identify possible partners we consider feasible and scientifically advantageous to stipulate agreement with;
- to establish an a-priory agreement regarding information sharing and publication policy with all the groups and facility directors involved in the hunting for the E.M. signals; the same should be done for the VIRGO-LIGO consortium, in relation with multi-frequency GW detection;
- to define a protocol for sending triggers to the LIGO-VIRGO consortium.

Outputs

- a list of partners that can be continuously updated in later phases, according to theoretical, technical and observational advancements and the work of WP 7.1;
- a proposed document that established the information sharing and publication policy;
- a specific protocol for sending triggers to the LIGO-VIRGO consortium.

Time-frame

- Phase 1: between month 1 and 10: in synergy with WPs in Section 7, review all possible *type* of partners (e.g. observatories, individuals, etc...) we would like to establish partnerships.
- Phase 2: between month 11 and 18: this WP will critically review all the existing strategy notably in place in the LIGO/VIRGO collaboration and in the γ-ray burst community to brainstorm on a possible document, tailored to LISA and the identified LISA partners (see phase 1). Establish partnerships and agree on protocol with them.
- Phase 3: between month 19 and 24: writing documents with outputs (see above).

Possible sub work-packages
I propose *not* to have sub-WPs for this WP.

Dependencies
This WP output is dependent on all WPs in Section 7.

7.6 Define data/communication protocols

Goals/Motivation

The aim of this WP is to identify the best strategy to release alerts for LISA triggers. In particular, the team should

- identify the minimum set of data from our analysis pipeline that must be communicated to enable a prompt follow-ups of our sources by E.M. devices;
- deliberate on how factually these data should be distributed between partners identified in WP 7.2: we should simultaneously and reasonably take into account both confidentiality and efficiency in the delivery process.

This WP focuses mainly on transient sources. VBs will be discussed separately in WP 7.6

Outputs

- A written document where the above goals are fully discussed. This document will be continuously updated also beyond phase A.

Time-frame This WP is closely related to WP dealing with alert distribution tools and low-latency and WP 7.1 (see "dependencies" below). It actually needs to build on the output of those WPs. As a consequence the actual work will start on month 12. In detail:

- First phase: between month 0 and month 11: latency time.

- Second phase: between month 12 and month 18 this WP members will collect information from other WPs and critically review practises concerning data circulations adopted by other observatories like LIGO/VIRGO and transient surveys.
- Third phase: between month 18 and month 24, the members will write the document with the proposed strategy.

Possible sub work-packages I propose *not* to have sub-WPs for this WP.

Dependencies

- all the WPs in Section 7
- WP described in Subsection 2.5
- all WPs in Section 4

7.7 Technical outreach and communication

Goals/Motivation

- Output gravitational wave data that is consistent and useable with EM data sets and catalogs
- For SMBHs: early sky localization error boxes, and continuing evolved error box as SNR grows and localization becomes better. Include also estimates of distances, orbital periods to seed EM counterpart searches.
- For GBs: sky localization, orbital periods, distance (for chirping binaries) will aid in EM follow-up observations
- For astrophysical interpretation, *orbital ephemeris* is needed (orbital elements are useful, but most importantly the orbital phase as a function of time).

Outputs

- A reporting structure to inform the astronomical community about potentially interesting sources in the LISA data (SMBHBs still a year or more from merger; strong galactic binaries, etc).

Time-frame

- Begin to develop notions about the useful data products for interface with EM astronomers within the Consortium, but collaborative discussions should be had with end-users of this data well before final pipelines are developed (est. ~ 5 years before launch)

Possible sub work-packages

- Define reporting procedures for sources (since in most cases, they are persistent, not transient events). "Alerts" are not the correct idea; circulars see more likely (e.g. in the spirit of Minor Planet Circulars). What is the format, what is the information content, and how are they produces, distributed, and archived?
- Significant package associated with galactic binaries, deriving for the global fit extraction
- SMBHB localization update and report (this will emerge from the global fit, but EM partners may need updates at odd intervals – e.g. coincident with observing schedules)

Dependencies

- Dependencies on global fit.
- Communication infrastructure with astronomy community.

8 Interpretation, key-science projects

This group of WPs summarizes and ensures the delivery of the main scientific goals of the mission, building upon the tools provided by the WPs of the previous groups and the associated pipelines. In more detail, LISA will detect the continuous and nearly monochromatic signal from ultra-compact binaries known from EM surveys (WP 8.1), and will detect a large number of ultra-compact binaries as GW-only sources (WP 8.2), detected as a foreground and as individual sources. LISA will probe the galactic population of white dwarfs, neutron stars and stellar origin black holes in binaries in a unique way, providing the first census of close binaries, shedding light into the processes of star formation and evolution in (interacting) binaries. But, LISA will explore also the universe beyond our Milky Way Galaxy. The LISA observatory has as prime objective to shed light into the processes of formation of the first seed BHs, formed in primeval halos (WP 8.3), and upon which the MBHs, ubiquitous in todays galaxies, have grown. LISA will let us peer deep into the young universe, when the first dark matter halos collapse and the first stars are forming, providing exquisite information on the mass spectrum of the earliest MBHs. LISA will also explore the hierarchical build up of MBHs by detecting the loud signal of coalescing MBHs of about million suns across all cosmic ages to probe their concordant evolution with galaxies and their clustering (WP 8.4). These are the major science deliverables of the mission, associated to WPs 8.3 and 8.4, and require the building of astrophysically-motivated interpretation-pipelines to tell apart competing models for the evolution of these sources. Since MBH binaries may trigger or be associated to electromagnetic counterparts in radio or optical/X-ray, a separate WP about the interpretation of the latter is spelled out in WP 8.5. Other classes of sources that are expected to be detected during the course of the mission are EMRIs, IMRIs, IMBHs and SOBHs. For each of these classes, we will perform result-interpretation and population-analysis studies with GW-only data (WP 8.6, 8.8 and 8.9, respectively). Deviations away from the predictions of general relativity will be scoped out in WP 8.7. LISA coalescing binaries are also standard sirens. Estimation of cosmological parameters (via GW-only observations combined in a statistical fashion, or via GW+EM observations) will be the subject of a dedicated pipeline (WP 8.10). LISA can also detect a SGWB of astrophysical and/or cosmological origin (WP 8.11): the detection (or lack thereof) of a stochastic background allows to impose constraints on astrophysical population models, models of fundamental high energy physics and the very early universe. The development of a framework to robustly detect and characterize unexpected/unknown sources of GWs in the LISA band is envisaged in WP 8.12.

Table 8 summarises the work packages of this group.

8.1 Analysis of joint GW+EM observations of GBs (including VBs)

Of all the sources LISA is sensitive to, only the known ultra-compact binaries ("verification binaries", VBs) are expected to be well studied and characterized before launch, through electromagnetic observations. Additionally, after launch, more galactic binaries are expected to be discovered that can be simultaneously observed in both gravitational waves and electromagnetic telescopes.

Goals/Motivation

- The goal of joint analysis of the VBs is to maximize the understanding of *single sources*, to validate instrument performance and analysis, as well as produce archetypes for comparison against other systems where complete GW and EM observational are not jointly available.

Outputs

WP	Description	Priority
8.1	analysis of joint GW+EM observations of GBs (including VBs)	3
8.2	population studies of GW-only GBs	3
8.3	studies of seed black holes and BH formation mechanisms	3
8.4	studies of SMBHBs and connection to galaxy clustering	3
8.5	analysis of joint EM+GW SMBHB events	3
8.6	analysis of EMRI population	3
8.7	tests of GR and the nature of compact objects	3
8.8	analysis of IMBHBs and IMRIs	3
8.9	studies of SOBH populations	3
8.10	estimation of cosmological parameters	3
8.11	characterisation of backgrounds	3
8.12	analysis of detected unmodelled events	3

Table 8: Work packages on interpretation and key science questions.

- Gravitational-wave catalogue of known verification binaries
- Discovery catalogue of new joint GW+EM binaries

Time-frame Many of the items needed here are covered by the WP in Sec. 7.6, as the analysis naturally flows from those characterizations.

- EM survey catalogs (LSST, GAIA, etc) should be canvased on release for candidate VBs
- A well defined ephemeris (and best fit EM parameters) will be needed at launch, requiring dedicated observing campaigns to be carried out 2-4yr before launch. The larger the aperture required for observations, the longer the lead time.
- 2-4yr prior to launch, a model for a guest observer program for the joint EM+GW analysis needs to be implemented and tested with current best EM data.

Possible sub work-packages

- Maintenance of ”known verification binaries” parameters from best EM obervations.
- Model pipeline to take the current best known parameters from EM observations and model the emitted GW signals and their related uncertainties.
- VBs are already known. EM work in advance needs to be carried out to build the best EM priors before LISA analysis.
- New EM surveys (eg. GAIA, LSST) need to be analyzed for candidate binaries, as do continuing EM surveys (eg. ELM)
- Quality assessment of known VBs (prior to launch) and newly discovered GBs as time/phase references for other analyses (e.g. bridging gaps in the data). Data products from this will depend on the needs for the work packages that consider data gaps.

Dependencies

- The standard GB searches (WP 1.4) should recover the VBs, as well as any joint binaries.
- EM catalogs of known binaries, with sky location and best EM parameters.

8.2 Population studies of GW-only GBs

The LISA sample of ultra-compact binaries will be among the largest well characterized catalogs of these sources, representing a substantial probe of the underlying population. The data will include both the individual sources ("resolved binaries") as well as the limiting confusion foreground.

Goals/Motivation

- To identify distinct sub-populations within the LISA characterization of the GBs as a probe of the evolutionary processes that drive the characteristics of these systems.
- Identification will be made by clustering in parameter space for the resolved GBs, and in frequency-dependent shaping of the irreducible foreground by sub-populations that dominate in a give frequency range.

Outputs

- parameter distributions of resolved GBs by type.
- Spectral description of the irreducible GB foreground (spectral amplitude and index, as a function of frequency).

Time-frame

- The pipelines for this analysis will use as input the data products of the global analysis and noise characterization data streams. As such, development can proceed at a low-level to devise strategies, but should begin in earnest once an initial data product is available.

Possible sub work-packages

- Improved population modeling from EM samples (e.g. Gaia, LSST, ELM), as well as extrapolating from LIGO to LISA (e.g. inferring NS binary population from LIGO BNS rates).
- Characterization of the variance in modeling the galaxy based on current best inputs for population synthesis (this tool exists, and will be publicly deployed in 18 months).
- Consistency of the population with the catalogues of VBs and joint EM+GW binaries (WP 8.1).
- Cyclostationary analysis; this has been examined in an initial study but never demonstrated on mock data.

Dependencies

- The catalog of GBs from the global fit (WP 5.3).
- The residual spectrum and its parameters; this includes estimates from Sagnac channels (WP 8.10), as well as time evolution fo the residual spectrum ("cyclostationary" behaviour).

8.3 Studies of seed BHs and formation mechanisms of massive BHs

BHs of $10^{6-9}\,M_\odot$ (usually referred to as MBHs or SMBHs) are ubiquitous in today galaxy's spheroids, and evidence is growing that BHs of about $10^5\,M_\odot$ or less inhabit dense, nuclear star clusters at the centres of bulge-less, disc galaxies. Consensus is rising that MBHs have been emerging from a population of *seed BHs* of yet unknown origin. Then, they increased their mass through repeated episodes of gas accretion, tidal captures of stars and coalescences with other MBHs in concordance with the evolution of cosmic structures. Possible seed formation mechanisms comprise: *(i) primordial* BH (PBH) with masses ranging between 0.01 to $10^5\,M_\odot$ that may have formed in the very early universe from peaks in primordial curvature fluctations; *(ii) light seeds* of about $200 - 10^3\,M_\odot$, representing the endpoint of the collapse of massive metal-free stars or metal poor stars beyond the pair-instability gap that formed around $z \sim 20$, when the universe was only 180 Myr old; *(iii) heavy seeds* of about $10^4 - 10^5\,M_\odot$ that may have formed later in gas rich, yet metal poor dense star clusters through sequential collisions of massive (even pre-main-sequence) stars or collisions among fragmenting gas clumps, or through the monolithic collapse of a massive cloud in a UV illuminated dark matter halo. Pairing of BH seeds and formation of "binary seeds" is still poorly explored in cosmological simulations (due to resolution limitations). LISA will shed light into this process as LISA shall have the capability to detect the inspiral signal of binary BH coalescences with masses down to $\approx 10^3, M_\odot$ (as measured in the source frame) out to $z \sim 20$, and the inspiral, merger and ringdown of BH binaries with masses in the interval between $\sim 10^4\,M_\odot$ and $\sim 10^7\,M_\odot$, even beyond $z = 20$.

Goals/Motivation

- Identify the population of the earliest MBHs, coalescing before the epoch of cosmic reionization, and draw preliminary conclusions on the physical origin of the BH seeds at cosmic dawn.
- Infer the rates of the earliest BH seed coalescences, and draw preliminary conclusions on their occupation fraction in galactic halos.
- Carry a comprehensive investigation to identify unmodelled GW signals from colliding BH seeds at high redshifts.
- Investigating the existence of MBHs of about $10^6\,M_\odot$ at $z \sim 15-20$, and draw conclusions on the formation pathways of MBHs shining as bright QSOs at $z \sim 7$.
- Set limits on the time lapse between formation and coalescence of the observed MBH merger events to indirectly infer (yet unknown) rates of MBH sinking via gas-dynamical processes in high-redshift dark matter halos.
- Explore whether heavy BH seeds continue to merge in halos during the epoch of shining of the brightest QSOs (at $z \sim 7$) down to cosmic high noon (at $z \sim 2$, when the bulk of the star formation takes place in all the galaxies) to draw preliminary conclusions on channels of potential delayed/continuous formation, on the long-term growth of seeds, or/and delayed pairing.
- Investigation on new channels of formation, driven by the GW detections.

Outputs

- Catalogue of the earliest BH binary coalescence events in the intermediate mass interval between about 10^3 and $10^5\,M_\odot$.

- Publication describing the properties of the earliest merging BHs and implications of this finding on the evolution and dynamics of very massive stars in metal poor galaxy halos and dense star clusters.
- Publication(s) describing the properties of the earliest merging BHs and implications of this finding on models of monolithic collapse of gas clouds for the formation of heavy BH seeds.
- Publication(s) describing the properties of the earliest merging BHs and implications on the scenario invoking the formation of PBHs.
- Publication(s) comparing the mass spectrum of the earliest BHs with the low-redshift BHs in the corresponding mass range, as inferred from the observations of EMRIs.
- Publication(s) of the effect of weak lensing on the estimate of luminosity distance of these high redshift signals.
- Study of potential match in redshift space between EM signals from forming BH seed and GW signal from merging seeds.

Time-frame

- Simulations and studies on the formation of very massive stars in dense, gas rich clusters, $\sim$ 10y.
- Simulations of BH seed formation in metal poor massive clouds with improved astro-chemistry and radiative transport, $\sim$ 10y.
- Theoretical studies of mass segregation and possible runaway merger of SOBHs in dense (nuclear) star clusters with upgraded mass spectrum consistent with the new findings by the LVC, $\sim$ 5y.
- Cosmological simulations of halo-halo pairing with seed BHs and self-consistent treatment of BH pairing in the gas/stellar environment of a high-z galactic halo, $\sim$ 10y.
- Construction of moch catalogues of seed BHs, $\sim$ 5y.
- Development and testing of model selection pipelines to determine the properties of the earliest seed BHs. Use the output of the investigations described above to construct a parametric semianalytic model and use observed distributions to pin down the relevant parameters, $\sim$ 10y.
- Publication and Archive of LISA findings from L+1yr.

Possible sub work-packages

- sub WP1: theoretical studies on the dynamics of massive stars and/or SOBHs in metal-poor star clusters - expertise to find/develop within the Consortium.
- sub WP2: theoretical studies on the fragmentation of massive clouds and formation of heavy BH seeds in the high redshift universe - expertise to find/develop within the Consortium .
- sub WP3: theoretical studies on the formation, evolution and clustering of dark matter halos with in-house seed BHs - expertise to find/develop within the Consortium.
- sub WP4: study of the dynamics of seed BHs in halo-halo collisions at high redshifts.

- sub WP5: building a model selection pipeline to develop within the Consortium.
- sub WP6: understand the possible signature of primordial black holes in LISA sources of this type.

Dependencies

- Waveforms and codes for massive BH and intermediate mass BH coalescences (WP 1.3, WP1.5, WP 5.3, WP 5.6).
- Source catalogues (WP 6).
- Collaboration with scientists world-wide working on this rich field embracing themes in astrophysics and cosmology.

8.4 Studies of MBHBs and connection to galaxy clustering

In the aftermath of a collision between two galaxies, the MBHs residing at their centres sink via stellar and gas-dynamical processes, forming a close Keplerian binary fated to coalesce. If, according to the current paradigm of galaxy formation, dark matter halos grow through clustering, then MBH binary (MBHB) coalescences appear inevitable. In this perspective, MBHBs pin point the places where galaxy assemble and grow. LISA shall detect the coalescence of MBHB with SNR of fifty and higher, in the mass interval between $10^4\,M_\odot$ and a few $10^7\,M_\odot$, across all cosmic ages - from $z \sim 20$, through the epoch of the rare brightest QSOs at $z \sim 7$, down to cosmic noon when the star formation history and AGN activity have their maximum and the local universe.

Goals/Motivation

- Analyse the set of observed MBH coalescences to draw conclusions on their growth mechanisms through accretion and mergers with other MBHs from the observed masses, mass ratios and spin distributions.
- Study the distribution in redshifts of the MBHB coalescences to draw conclusions on possible delays between time of formation of the binary and time of coalescence, to learn on the dissipative mechanisms driving MBHs toward coalescence, and study the delay-time distributions, by matching the data with population synthesis models.
- Learn whether MBH mergers retain memory of large initial eccentricities that might indicate interactions with a third MBH and activation of Kozai-Lidov resonances or chaotic three body encounters.
- From the rate of MBHB coalescences, draw conclusions on the merger rate of galaxies, and on the relative contribution between major and minor mergers, from the distribution of their mass ratios.
- Compare the population of merging MBHs with the MBHs sites of EMRI events, and draw conclusions on the population of "quiescent" MBHs versus the population of the "interacting" MBHs.
- From the spin distribution of the merging MBHs, draw conclusions on the nature of the accretion process (e.g., whether the accretion flow is preferential coherent or chaotic).
- In case of a conjunct EM + GW detection, correlate the properties of the EM signal with the properties of the two merging MBH, and compare with those that did not show a counterpart, to draw conclusions on the environment.

Outputs

- Publication(s) describing the mass distribution of observed MBH coalescences and its implications for the overall growth of MBHs.
- Publication(s) comparing the masses of the MBH in the LISA data stream with those resulting from EM observations, possibly at selected epoch, within narrow redshifts bins.
- Publication(s) describing the distribution in mass ratios of the observed coalescing binaries to infer properties on the underling host galaxies
- Publication(s) describing the spin distributions of the observed merging binaries and discuss the implications of the nature of the accretion flows.
- Publication(s) aimed at comparing the distribution of spins of the observed coalescing MBHBs with the spin of isolated SMBH seen in nearby galaxies through the K-iron lines, and with the spins of MBHs hosting EMRIs (i.e. quiescent MBHs).
- Publication(s) describing how the parameters of the observed GW events (masses, mass ratio, spin orientations and residual eccentricity) fit with the grand scenario of galactic collisions resulting from state-of-the-art simulations.
- After generating a mock data set collecting the spins of accreting MBHs, both in magnitude and direction, use observations to address the inverse problem (i.e., model inference from the data).

Time-frame

- Development of novel population synthesis models for MBHBs using cosmological simulations and improved sub-grid physics. Carry out periodic upgrades and refinement of the models, to update recipes in relation to advances in knowledge on how galaxies grow and evolve from the then-current observations, e.g. with ALMA and JWST, L+0y
- Construction of an archive of ”galaxy mergers” to instruct on the search and identification of the GW signals, $\sim$ 3y
- Development of the pipeline to address the inverse problem, $\sim$ 5y

Possible sub work-packages

- sub WP1: Dynamics of merging MBHs in high redshift galaxies with initial conditions consistent with updated cosmological simulations.
- sub WP2: Detailed studies on the MBH spin evolution in their interaction with a circum-nuclear disc prior to merge, and analysis of their waveforms.
- sub WP3: Studies of the dynamics of MBHs in low-redshift, gas poor galaxies to shed light on the interactions between MBHs and stars.

Dependencies

- Waveform and PE codes for MBHBs (WP 1.3,WP 5.3).
- Source catalogues (WP 6).
- Population synthesis models and numerical studies of mergers.
- Collaboration with the communities working on AGNs and on galaxy formation.

8.5 Analysis of joint EM+GW MBHB events

LISA will observe MBHB mergers along the cosmic history, providing key information about their assembly in connection with their galactic host (see WP 8.4). MBHBs form in the centre of galaxy merger remnants, surrounded by an environment that is expected to be very dense and (especially at high redshifts) gas rich. Detection of EM counterparts would provide invaluable extra information, extending the realm of multi-messenger astronomy to the high redshift Universe. Exploitation of joint GW and EM information is therefore key for a fulfil the full LISA potential.

Goals/Motivation

- Identify MBHB hosts to study their physical environment in connection with galaxy clustering and structure formation up to high redshifts.
- Connect the nature of EM emission with the parameters of the merging MBHs. E.g. are the MBHs associated with EM counterparts preferentially highly spinning? Are their spin aligned?
- Gain insight on the environment in the immediate vicinity of merging MBHs and constrain the underlying physics of the emission. E.g. what is the SED of the emitted light? Is there evidence of jets? Can we extract the properties of the accretion flow?
- Compare findings with proposed models (e.g. full GRMHD simulations) and develop a coherent picture of the MBHB matter interaction in the inspiral-merger-post merger phase.

Outputs This working package should operate with theoretical endeavours aiming at modelling MBHBs in gas rich environments. Although some of the output might not directly controlled by the Consortium, the Consortium should support the efforts in this direction.

- Publication(s) of a thorough survey and categorization of the physical interaction between MBHBs and their environment, with particular emphasis on: i) the connection to the MBHB parameters and ii) the associated EM counterparts.
- Detailed HD, MHD and GRMHD simulations of merging MBHBs. The HD, MHD simulations at larger scales will provide the initial conditions for the GRMHD at smaller scales (possibly outside the Consortium).
- Simulations of a range of joint EM-GW signals and test of multimessenger potential on mock LISA (and EM) data. Publication of the results.
- Theoretical framework to connect EM emission and GW source parameters. Delivery a pipeline for astrophysical inference.
- Publication of joint GW-EM observations of individual LISA events (requires MOUs with partner EM observatories).
- Archive of all joint GW-EM detections with all the relevant data (requires MOUs with partner EM observatories).

Time-frame

- Theoretical studies of EM counterparts $\sim$ 3y.
- Detailed simulations of MBHB-gas systems $\sim$ 10y.

- First simulations and tests on mock data $\sim$ 3y.
- Theoretical framework for astrophysical inference from joint observations L+0y.
- Publication and archive L+1y.

Possible sub work-packages

- sub WP1: theoretical and numerical studies of EM counterparts.
- sub WP2: simulated detections of GW+EM signals for a range of future observatories.
- sub WP3: reverse engineering problem: inference on astrophysical models for the accretion flow and the MBHB-environment interaction from observations.
- sub WP4: archive construction and maintenance.

Dependencies

- Tight connection with WP 7.1 and WP 7.5 for the identification of EM counterparts and the development of codes for mock simulations of GW signals plus EM counterpart.
- Waveforms and PE codes for MBHBs (WP 1.2, WP 5.3).
- Source catalogues (WP 6).
- MOU with partner observatories and protocols (WP 7.2, WP 7.3).
- Studies of the MBHB population along the cosmic history (WP 8.3, WP 8.4).
- Collaboration with analytical and numerical modellers of MBHBs in gas rich environment.

8.6 Analysis of EMRI population

Goals/Motivation

- Analyse the set of observed EMRIs to draw preliminary conclusions about the astrophysical population, hence completing SI3.1 from the LISA proposal.
- Infer the rate of EMRI events, and its dependence on MBH parameters. Infer the distribution of EMRI orbital parameters (e.g., inclination, eccentricity). Draw conclusions about dynamical models of galacto-centric stellar clusters.
- Characterise the distribution of MBH masses and spins in the observed EMRI population. Draw conclusions about the astrophysical MBH population in the LISA range. Compare this 'quiescent' population to 'active' population observed in MBHBs.
- Characterise the mass function of compact objects in EMRIs. Draw conclusions about the stellar IMF in clusters and the efficiency of mass segregation.
- Look for evidence for primordial black hole of black holes involved in EMRIs.

Outputs

- Publication(s) describing the observed EMRI population and its implications for stellar clusters.
- Publication(s) describing the properties of the stellar compact objects observed participating in EMRIs and the implications for the cluster IMF (possibly combined with the above).
- Publication(s) describing the properties of the low-redshift MBH population (possibly combine with paper on MBH population observed in mergers, WP 8.4).

Time-frame

- Development of the tools for the main analysis is low-priority, as this will come after other data products. Some preliminary theoretical studies have already been done.
- Understanding if the EMRI population can tell us something more of astrophysical significance, e.g., if some subset of the observed EMRIs could be of primordial black hole origin, would be a higher priority, as it might impact on how the EMRI search and population analysis is finally done.
- Some methodological and pipeline development required by launch.
- Main effort connected to analysis and paper-writing L+5yr.

Possible sub work-packages

- Properties of the stellar population in galactic clusters inferred from the observed EMRI population.
- Dynamics in galactic clusters inferred from the observed EMRI population.
- Properties of the quiescent SMBH population at low-redshifts inferred from observed EMRIs.

Dependencies

- EMRI search and characterisation (WP 5.4). The primary data product that this WP will use is the catalogue of EMRI events generated in 5.4. These will in turn rely on EMRI waveform models developed under WP 1.2.
- Instrument performance monitoring (WP 2.8, 2.10, 2.11, 4.3, 5.10). Inference on populations relies crucially on understanding selection effects, which can only be done with an understanding of instrument performance.
- If EMRIs could have primordial black hole origin, this WP will depend on WP 8.11 to ensure consistent treatment of any (non-)detection of a PBH background.

8.7 Tests of GR and the nature of compact objects

LISA will provide tests of the nature of compact objects and of the behavior of gravity with unprececented accuracy in the strong-field, highly relativistic regime in which deviations away from GR are more likely to appear. To this purpose, a dedicated pipeline needs to be built in order to allow performing these tests and discriminate these effects from enviromental ones.

Goals/Motivation

- Test GR in the strong-field, highly relativistic regime.
- Test the nature of compact objects, and particularly GR's "no-hair" theorem.
- Detect and disentangle the effects of the interaction with matter ("environmental effects").

Outputs

- Waveforms for all model signals, including parametrized deviations from GR (in the generation and propagation of the signal), environmental effects, exotic near-horizon physics and the possible presence of massive degrees of freedom.
- Data-analysis pipeline yielding Bayes factors for the hypotheses of: i) deviations from GR, ii) environmental effects and iii) near-horizon physics/massive degrees of freedom, as well as estimates of the corresponding parameters, based on individual/stacked sources and on measurements of the stochastic background.
- Framework to translate parametrized constraints into specific theories/physical models.

Time-frame

- Low priority, because the final version of the pipeline will heavily depend on our future understanding of beyond-GR effects from ground-based detectors, as well as on future theoretical developments.
- Preliminary pipeline in 5yr, final pipeline within 10yr.

Possible sub work-packages

- Search for deviations from GR in the inspiral.
- Searches for deviations from GR in the merger/ringdown.
- Propagation effects.
- Searches for environmental effects.
- Searches for near-horizon exotic physics.
- Data-analysis pipeline.
- Interpretative framework.

Dependencies

- Waveform modelling (WPs 1).
- Data analysis tools (WPs 2).
- Multi-band GW analysis (WP 7.4).
- Collaboration with the larger community of GW theorists.

8.8 Analysis of IMBHBs and IMRIs

The existence of IMBHs in star and globular clusters, as well as in the nuclei of dwarf galaxies is subject of active debate. LISA will shed light on this open issue through its capability of detecting both IMBHBs and IMRIs – binaries comprising an IMBH and a SBH or NS or WD, i.e. with mass ratio $q \sim 10^{-3} - 10^{-2}$ – up to moderate redshifts.

Goals/Motivation

- Identify IMBHBs merging at low redshift and try to assess their nature. Are they late merging seed BHs or IMBHs harbored in globular/star clusters?
- Analyse the set of low-redshift IMRIs events and draw conclusions on their formation processes in dense star clusters.
- Characterise the distribution of mass ratios and spins to draw conclusions on the dynamics in the core of clusters.
- Compare this populations with the EMRI's population as well as high redshift seed BH population.
- Infer the yet unconstrained IMRI's rates.

Outputs

- Theoretical models for IMBHs and IMRI formation in star clusters.
- Publication(s) describing the low redshift IMBHB population and its implication for stellar clusters.
- Publication(s) describing the observed IMRI population and its implication for stellar clusters.
- Publication(s) comparing the low redshift IMBHB/IMRI with EMRIs and the high redshift seed BHs, to understand whether the former belong to a different population.

Time-frame

- Study the formation channel for IMBHBs and IMRIs in star clusters $\sim$ 3y.
- Pipeline for model inference from IMRI/IMBHB observations $\sim$ 5y.
- IMRI/IMBH catalogs, from L+1y.

Possible sub work-packages

- sub WP1: theoretical studies of IMBHB and IMRI formation in star clusters and dwarf galactic nucley via dedicated N-body simulations and characterization of the expected population.
- sub WP2: inference from observations. Constrain the astrophysics of IMBH in stellar clusters including possibly formation efficiency, mass funcion and occupation fraction.

Dependencies

- Waveforms for IMRIs (WP 1.5) and comparable mass BHs (WP 1.3).
- Detection pipelines for IMRIs (WP 5.6) and comparable mass BHs (WP 5.3).
- Source catalogues (WP 6).

8.9 Studies of SOBH populations

LISA will observe several SOBH in the relatively early inspiral phase, providing complementary information to ground based observations, in particular measuring eccentricity and possibly environmental influences and spins. Moreover, LISA will observe the stochastic GWB produced by the large populations of unresolved events at high redshift. The information gathered by LISA (combined with that provided by ground based detectors) will provide deeper insights in the astrophysical origin and dynamics of SOBHBs, and will also possibly allow to thest the hypothesis of their origin in the primordial universe.

Goals/Motivation

- Keep up to date with ground based observations (observation-driven catalogues) and produce mock catalogues of SOBH accordingly to be injected in LISA mock data streams.
- Study possible signatures of different formation scenarios (i.e. eccentricity distribution, spin distribution, source acceleration, etc) and environmental effects (i.e. dirty SOBHs, perturbation from a third object) and their imprint on the expected waveforms.
- Develop appropriate analysis tools for the identification of astrophysically relevant features in the GW waveforms.
- Develop tools for inference from data and model selection and test them on recovered signals from LISA mock data injection.
- Pin down SOBHB formation channels, their relative contribution to the observed population and understand the connection between stellar evolution, dynamics in dense environments and SOBHB populations (Final key-science goal).

Outputs

- Catalogues for mock data for injection.
- Publication of a thorough analysis of possible SOBHB signatures in the LISA band and their connection to SOBH astrophysics and dynamics.
- Astrophysical population inference pipelines.
- Publication of results from the analysis of LISA events, including dynamical properties of SOBHBs, contribution to difference formation channels, connection with other populations of known binaries of compact objects (NSB, BH-NS), etc.

Time-frame

- Mock catalogues will be produced periodically based on ground based updates. The first catalogues have already been shared in the LISA MDC git.
- Study and publication of astrophysical and dynamical mechanisms leaving possible signatures in the waveform ∼2 yr.
- Development and tests of astrophysical population inference pipelines ∼3-5 yr.
- Final astrophysical inference pipelines required by L+1yr.

Possible sub work-packages

- Sub WP 1: astrophysics and dynamics of SOBHs.
- Sub WP 2: population inference from data.

Dependencies

- SOBHB waveforms (WP 1.6).
- Individual and global source identification codes (WP 5).
- Source catalogues (WP 6.1, WP 6.2. WP 6.3).
- External collaboration with population synthesis modellers and with population modellers investigating the PBH hypothesis, especially for astrophysical and cosmological inference from data.

8.10 Estimation of cosmological parameters

GWs from compact binaries provide a direct measurement of the luminosity distance of the emitting binary. For sources with SNR larger than 8, if the two polarisations can be distinguished, d_L can be determined with good accuracy: the main limiting factors for the accuracy often being the intrinsic peculiar velocity and weak lensing. The redshift of the source can be inferred either via the detection of an electromagnetic (EM) counterpart, or statistically by assigning a random host galaxy in the GW "sky error box" to the detected binary. Once both the redshift and the luminosity distance of a catalogue of sources are known, it is possible to fit the luminosity distance-redshift relation of the background Friedmann-Lemaitre-Robertson-Walker metric to infer the cosmological parameters. The latter include the present value of the Hubble constant H_0, the matter density parameter Ω_M, the cosmological constant/dark energy parameter Ω_Λ; depending on the modelling of the dark energy component, more parameters can be relevant, e.g. the equation of state of dark energy today, w_0, its variation with the scale factor, w_a, etc.

Goals/Motivation

- Probe the nature of the acceleration of the universe using GW instead of EM observations: the advantages are different systematic effects and the possibility of testing redshifts higher than those of usual cosmological probes like SNIa and galaxy catalogues.
- If the redshift is inferred via statistical techniques, the measurement of the luminosity distance-redshift relation can be done fully independently of EM observations, therefore offering a completely new approach to understand the nature of the mysterious cosmological constant/dark energy.
- Given the precision with which it can determine the Hubble constant, LISA will provide a new measurement of H_0 that can shed light on the tension between the values measured by CMB probes and by local-universe observations (like cepheids and SNIa).

Outputs

- A catalogue of MBHBs detected by LISA with SNR > 8 and sky localisation $\Delta\Omega < 10\,\mathrm{deg}^2$ both using inspiral only and inspiral+merger and ringdown, created considering the state of the art MBHB formation models. An estimate of the EM signal should be inferred for the MBHBs of this catalogues using state of the art simulations of the counterpart emission in radio and/or optical, with the aim of constructing a catalogue of standard sirens composed by: MBHBs with SNR>8 in LISA, sky localisation $\Delta\Omega < 10\mathrm{deg}^2$, and the possibility of determining their redshift via EM counterparts detectable by EM telescopes operating at LISA time (SKA, LSST, ELT...).

- For the above mentioned catalogue, a code to extract the accuracy on the measurement of the cosmological parameters for several cosmological models: flat universe, cosmological constant, dark energy with constant equation of state, dark energy with varying equation of state, early dark energy, interacting dark energy, possibly modified gravity scenarios and so on [with publication of results].
- A catalogue of MBHBs detected by LISA with SNR > 8, sky localisation $\Delta\Omega < 10\text{deg}^2$, and error on the luminosity distance $< 5\%$, both using inspiral only and inspiral+merger and ringdown, created considering the state of the art MBHB formation models, to be used with statistical techniques for the redshift identification.
- A catalogue of SOBHBs detected by LISA with SNR > 8, sky localisation $\Delta\Omega < 10\text{deg}^2$, and error on the luminosity distance $< 20\%$, created considering the latest information on the population, to be used with statistical techniques for the redshift identification.
- A catalogue of EMRIs detected by LISA with SNR > 10, sky localisation $\Delta\Omega < 10\text{deg}^2$, and error on the luminosity distance $< 10\%$, created considering the latest information on the rates of detection and the waveforms, to be used with statistical techniques for the redshift identification.
- A code to extract, from the above mentioned catalogues, the accuracy on the measurement of the cosmological parameters for several cosmological models (mentioned above) using the statistical techniques for the redshift identification that assign a redshift to the GW event based on simulated or observed galaxy catalogues. The code should work for single-type source catalogues and for the combination of all the above mentioned catalogues, with publication of the results.

Time-frame

- This WP is low priority. Some forecasts have already been carried out, other forecasts with updated catalogues can be done and published on a time-frame of 5-10 y.
- The analyses connected to this WP should be kept up to date on a yearly basis in what concerns products of other packages such as population models, waveforms and sensitivity of the instrument, and in what concerns cosmological constraints and cosmological models (from other cosmological probes)
- The full analysis with data and the paper writing can be carried out on a schedule L+5y.

Possible sub work-packages No sub work-packages have yet been identified for this WP.

Dependencies

- WP 1.2, 1.3, 1.6 for the population models and the generation of catalogues.
- WP 2.6 for delensing.
- WP 8.2 for the modelisation and detection of the EM counterparts to MBHBs.
- WP 8.9 for MBHBs formation models.
- WP 7.1 and 7.2 for the MoU guaranteeing the detection of the EM counterparts and access to galaxy catalogues data to be used in the statistical techniques for the redshift identification and for delensing.

- Results on cosmological parameter values from other cosmological probes (SNIa, CMB, BAO, galaxy catalogues, redshift space distortion, Lyman alpha...)
- Results and updates on the viability of dark energy/modified gravity models from the theoretical cosmology community.

8.11 Characterisation of backgrounds

The superposition of many unresolved sources appears as a stochastic signal in the detector. A stochastic gravitational wave background (SGWB) can be of astrophysical origin: in the LISA band we expect at least three types of astrophysical SGWB, from unresolved GBs, unresolved SOBHBs (which could also be of primordial origin), and unresolved NSBs. Astrophysical SGWB from unresolved EMRIs is also possible. SGWB can also be of cosmological origin, generated by a variety of processes operating in the very early universe. Sources susceptible of emitting in the LISA band include: primordial first order phase transitions at the electroweak scale and beyond, topological defects, non-standard models of inflation, PBHs with masses of the order of $10^{-12}\,M_{\odot}$.

Goals/Motivation

- SGWB of astrophysical origin: a measurement or upper bound would allow to understand and confirm population models, possibly disentangling the PBH hypothesis from the SO hypothesis for BHB of a few tents of solar masses.
- SGWB of cosmological origin: its discovery would provide information on the fundamental physics model describing the early universe beyond the standard model of particle physics, complementary to the one gathered from CMB and collider searches (for example, the presence of new degrees of freedom at the TeV scale, of extra dimensions or sectors uncoupled to the standard model, the dynamics and particle content during inflation, the occurrence of phase transitions leading to topological defects and so on).
- An upper bound on the amplitude of a SGWB allows to constrain models of high energy physics at and beyond the TeV scale, inflation beyond the single-field slow roll scenario, and the presence of PBH with masses of the order of $10^{-12}\,M_{\odot}$.

Outputs

- Estimation of the instrument noise and its time evolution (via ground tests, null-stream TDI channels as Sagnac, and other techniques available).
- the amplitude and spectral shape of the highest-SNR SGWB: the measurement of the shape of the signal is fundamental to distinguish its origin. Data analysis packages extracting a given number of parameters characterising the SGWB should be provided, as well as binned data points of the residual SGWB once all sources (including astrophysical SGWB) are substracted, for the community to fit the biggest variety of models.
- Depending on the SNR and measurement accuracy of the highest-SNR SGWB parameters, its composition should be assessed to understand whether it is originated by one or more sources. In the second case, packages to separate the signal in its principal components should be provided, and each component should be characterised in terms of amplitude and spectral shape to infer its origin.
- standard classes of SGWB models detectable in the LISA band should be defined in order to search for them in the data, in complementarity to blind searches

- Depending on the SNR and measurement accuracy of the highest-SNR SGWB, the presence of anisotropy, non-gaussianity and chirality will be assessed.
- Characterisation of the influence of a SGWB on the ability of detecting other sources. As the level of the cosmological SGWB is difficult to predict, this characterisation should be performed based on updated predictions of the astrophysical component. At present we expect $30 \lesssim \mathrm{SNR} \lesssim 300$ for the background from SOBHBs.
- Given a null detection, compute constraints on theoretical models. Estimate dependence of background amplitude on population parameters under a number of different scenarios. This should include backgrounds generated by primordial black holes and by cusps or kinks on cosmic strings.

Time-frame

- The urgency of characterising/removing the SGWB depends on its level, as it may influence the detection of other GW sources. The level of a cosmological background is difficult to predict, therefore the time-frame should be established based on predictions of the astrophysical background level, which should be kept up to date (cf WP 8.9). Depending on the accuracy level of the SGWB detection pipelines, we establish for now the following time-frame: 2y for basic analyses, 5y for accurate ones, 10y for final ones. Some forecasts have already been done for LISA.
- The analyses connected to this WP should be kept up to date on a yearly basis in what concerns products of other packages such as population models and sensitivity of the instrument, and in what concerns predictions for the cosmological stochastic background.
- The full analysis with data and the paper writing can be carried out on a schedule L+5y.

Possible sub work-packages

- For each cosmological SGWB source, accurately predict the amplitude, spectral shapes and other possible features of the corresponding SWGB, depending on the source characteristics.
- Develop pipelines to infer SGWB parameters, principal components, and other characteristics (non-gaussianity, isotropy...).
- Understand the constraining power of a null detection.

Dependencies

- WPs about noise characterisation such as WP 2.7-WP 2.10.
- WPs about sources that can be hidden by the known astrophysical SGWB: SOBHBs (WP 8.8), GBs (WP 8.5), EMRIs (WP 8.6), cosmic strings (WP 1.7).
- Take into account the potential of future Earth-based detectors such as the Einstein Telescope and Cosmic explorer to reduce, by accurate detections of resolved events, the level of the astrophysical SGWB.
- Take into account possible predictions/detections of the SGWB from future Earth-based detectors and from new discoveries at the Large Hadron Collider.
- Constraints on backgrounds generated by primordial black hole or cosmic strings should take into account the observed population of black holes (WP 8.3, 8.6) and any detected modelled transients consistent with bursts from cosmic string cusps (WP 5.9).

8.12 Analysis of detected unmodelled events

Deciphering unexpected and unmodelled signals will be very challenging, but also very rewarding as discovering entirely new phenomena will have high scientific impact. Here the analysis will be akin to detective work, gathering clues, forming hypotheses and devising new tests. Astronomy provides many examples of such detective work, including the discovery of active galactic nuclei, pulsars and gamma ray bursts. Finding definite evidence to support one model over another can sometimes take decades. The origin of short-hard gamma ray bursts was only completely settled recently when LIGO/Virgo detected a neutron star merger in coincidence with Fermi detecting a gamma ray burst. Other mysteries in astronomy remain unresolved to this day, such as the origin of fast radio bursts. While it is very difficult to anticipate the unexpected, we can ensure that a wide range of diagnostic tools are in place to help piece together the puzzle.

Goals/Motivation

- Discovering a new astrophysical phenomena may be the greatest scientific contribution from LISA.
- Detailed characterization of unmodelled signals can provide the clues needed to reveal the origin of the signals.

Outputs

- Robust waveform reconstructions with uncertainties.
- Tools to characterize the time/frequency content of the signals.
- Rapid localization to assist the search for electromagnetic counterparts.
- A test suite of hypothetical signals to check the analysis.

Time-frame

- Wavelet based waveform reconstruction techniques already used by LIGO/Virgo. Can be adapted for use with LISA data in $\sim$ 2y.
- Tools to characterize the time/frequency content of the signals already used by LIGO/Virgo. Can be adapted for use with LISA data in $\sim$ 2y.
- Simple examples of test signals can be developed quickly. More realistic models may take years to develop.

Possible sub work-packages

- Wavelet based waveform reconstruction.
- Alternative method for waveform reconstruction.
- Sky locations of short duration bursts.
- Tools to generate test signals.

Dependencies

- Detection and parameterisation of unmodelled signals (WP 5.5).
- LISA Instrument Simulator (WP 3.1).
- TDI models for noise transients (WP 3.3).
- TDI models for generic signals (WP 3.3).

Acronyms and Glossary

Acronyms

BH Black Hole

DCC Data Computing Center: physical center with computing and storage facilities

DPC Data Processing Center (Consortium)

CO Compact Object

EMRI Extreme Mass Ratio Inspiral

EM Electromagnetic

ESA European Space Agency

GB Galactic Binary

GR General Relativity

GS Ground Segment

GW Gravitational Wave

LISA Laser Interferometer Space Antenna

LDC LISA Data Challenge

MOC Mission Operation Center

PCT Payload Coordination Team

SCT Science Coordination Team

SMBH Supermassive Black Hole

SMBHB Supermassive Black Hole Binary

SOBH Stellar Origin Black Hole

SOBHB Stellar Origin Black Hole Binary

SGWB Stochastic Gravitational Wave Background

SOC Science Operation Center (ESA)

VP Verification Binary

Glossary