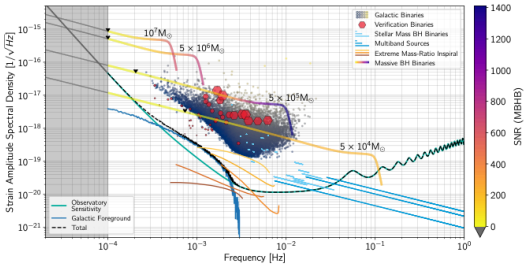
The LISA Science Ground Segment

Tyson B. Littenberg, NASA MSFC
Michael L. Katz, NASA MSFC

INTRODUCTION

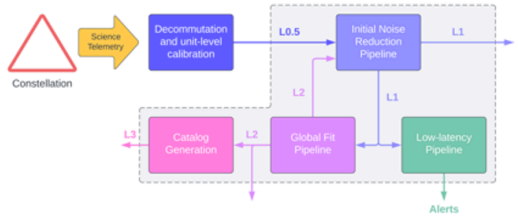
The mHz band of the gravitational wave (GW) spectrum contains an unprecedented abundance of galactic, extra-galactic, and cosmological sources. The Laser Interferometer Space Antenna (LISA) will survey the mHz GW band and provide unique observational constraints on the formation and evolution of compact binaries in the Milky Way, the origin and growth of massive black holes throughout cosmic history, the dynamics of dense stellar environments in galactic nuclei, the fundamental nature of gravity and black holes, and more. The richness of the LISA data comes at the price of a more complicated analysis framework than is required for currently operating GW observatories.



Caption: Power spectrum of expected LISA sources compared to the design sensitivity curve (black line). Most of these signals are long-lived in the LISA band, overlapping one another. Analysis algorithms must jointly fit to the full ensemble of GW sources in the LISA band to avoid biases or misclassifications when characterizing sources.

NASA is developing end-to-end data analysis algorithms which will be part of the overall mission SGS. The resulting pipelines will be used to identify and characterize individual GW sources. Tools to help scientists utilize this information to conduct a wide array of exciting scientific studies will also be developed as part of the NASA SGS.

LISA DATA PROCESSING



Caption: Flow chart of main data processing elements for LISA SGS

NASA will receive L0.5 data from the LISA Science Operation Scenter and produce L1, L2, and L3 data which will be returned to the SOC for archiving and distribution, including through a mirror at the NASA LISA Science Center.

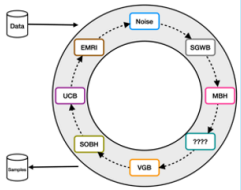
Data processing is divided into four main elements: The Initial Noise Reduction pipeline produces cleaned data used for GW searches. The Low-latency Pipeline identifies transient sources in real time to produce alerts form multimessenger observing partners. The Global Fit Pipeline performs the joint analysis of all GW sources in the data. The Global Fit output (L2 data) is a statistical representation of models consistent with the data, which are consolidated into a source catalog (L3 data) by the Catalog Generation pipelines.

THE GLOBAL FIT ALGORITHM

LISA's L1-L2 processing is a particular challenge due to the high source density in the data. MSFC is pioneering the development of these so-called "Global Fit" pipelines.

A key element for the pipelines is to explore a variable dimension parameter space and account for correlations between individual sources and to classify different source types in the data

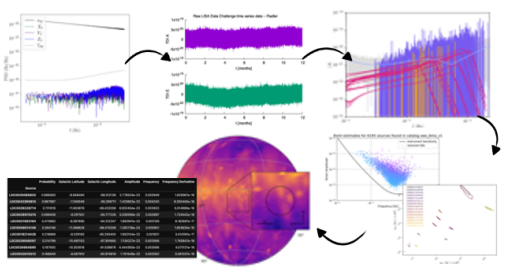
The Global Fit pipelines use Blocked Gibbs sampling algorithms to efficiently explore the large (>10^5) dimension parameter space.



Caption: Conceptual algorithm design for blocked Gibbs sampling global fit

In the Blocked Gibbs scheme groups of parameters (blocks) known to be correlated are updated jointly, while the rest of the parameters are held fixed. The algorithm then moves between blocks of parameters. Blocking schemes are hierarchical, with the highest level being different source types.

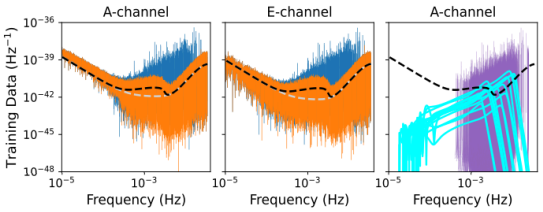
PROTOTYPE PIPELINES



Each data processing stage is in the prototyping phase at MSFC. Clockwise from top left: Power spectrum of laser frequency noise before (black solid line) and after (colored lines) L0.5-L1 processing; time series of L1 data containing population of UCB and MBH signals; reconstructed UCB, VGB, and MBH waveforms (purple, orange, magenta) extracted from time series data using L1-L2 pipeline; Point estimates in frequency-amplitude space of UCB catalog, and component mass posterior distributions for MBH population after L2-L3 processing; example user tool to query catalog for sources consistent with a region on the sky.

HARDWARE ACCELERATION

MSFC has parallel prototyping of global fit algorithms to study performance and cost trades, including a GPU-optimized pipeline which will significantly reduce energy costs for the L1-L2 data processing. The GPU pipeline has demonstrated comparable performance to the original CPU-based codes on the available simulated data sets.



Caption: Residuals (left and center) and reconstructed waveforms from GPU-optimized pipeline