Internship Proposal

Generative AI Models for Denoising and Glitch Mitigation in LISA Data

Laboratoire des 2 Infinis – Toulouse (L2IT, CNRS / Université de Toulouse)

Context

The Laser Interferometer Space Antenna (LISA) will open a new observational window in the millihertz gravitational-wave (GW) band, allowing the detection of signals from massive black-hole binaries (MBHBs) and thousands of galactic binaries. However, LISA's data will also contain non-Gaussian noise artifacts ('glitches') that can obscure or distort astrophysical signals. Traditional denoising and glitch mitigation techniques—largely designed for ground-based detectors—are not well adapted to the continuous and overlapping signal regime expected for LISA.

Recent advances in diffusion-based generative models—notably the Denoising Diffusion Restoration Model (DDRM) and the Diffiner architecture—have demonstrated remarkable performance in removing structured noise in complex, real-world signals (e.g., speech, audio, images). Adapting these methods to gravitational-wave strain data represents a promising and unexplored direction.

Objectives

The intern will explore the application of diffusion-based generative models to denoise and restore LISA strain data affected by instrumental glitches and overlapping MBHB signals. The project will involve both methodological development and experimental validation on simulated datasets.

- Specific goals:
  - Generate and preprocess simulated LISA strain data containing MBHB signals and glitch injections, using existing LISA Data Challenge datasets as a baseline.
  - Implement and adapt the Diffiner and DDRM architectures for one-dimensional (time-series) strain data, exploring conditioning strategies (e.g., frequency domain embeddings, time-frequency representations).
  - Train models on synthetic datasets and evaluate their ability to reconstruct clean MBHB waveforms.
  - Quantify denoising quality (e.g., SNR improvement, residual power spectra) and assess robustness to varying glitch morphologies and signal-to-noise ratios.

Expected Outcomes

- A prototype diffusion-based denoising model tailored for LISA-like GW data.
- Quantitative evaluation of the model's ability to separate astrophysical signals from glitches.
- (Optional) A short research note or contribution to a future LISA Data Challenge.

Skills and Tools

- Background in machine learning (Python, PyTorch or JAX)
- Familiarity with time-series/signal processing.
- Basic knowledge of gravitational-wave physics is a plus (training provided).
- Programming: Python, Git, Linux environments.

Internship Duration

4–6 months (Master 2 or engineering school level)

Supervision

The internship will be conducted at L2IT (Toulouse). The student will work in collaboration with scientists involved in LISA data analysis and machine learning for physics.

Contact

Antsa Rasamoela, antsa.rasamoela@l2it.in2p3.fr

References

- Sawata, R., et al. (2023). Diffiner: A versatile diffusion-based generative refiner for speech enhancement. Interspeech 2023.
- Kawar, B., et al. (2022). Denoising Diffusion Restoration Models. NeurIPS 35.
- Colpi, M., et al. (2024). LISA Definition Study Report. arXiv:2402.07571.
- Cuoco, E., et al. (2025). Applications of machine learning in gravitational-wave research. Living Reviews in Relativity, 28(2).