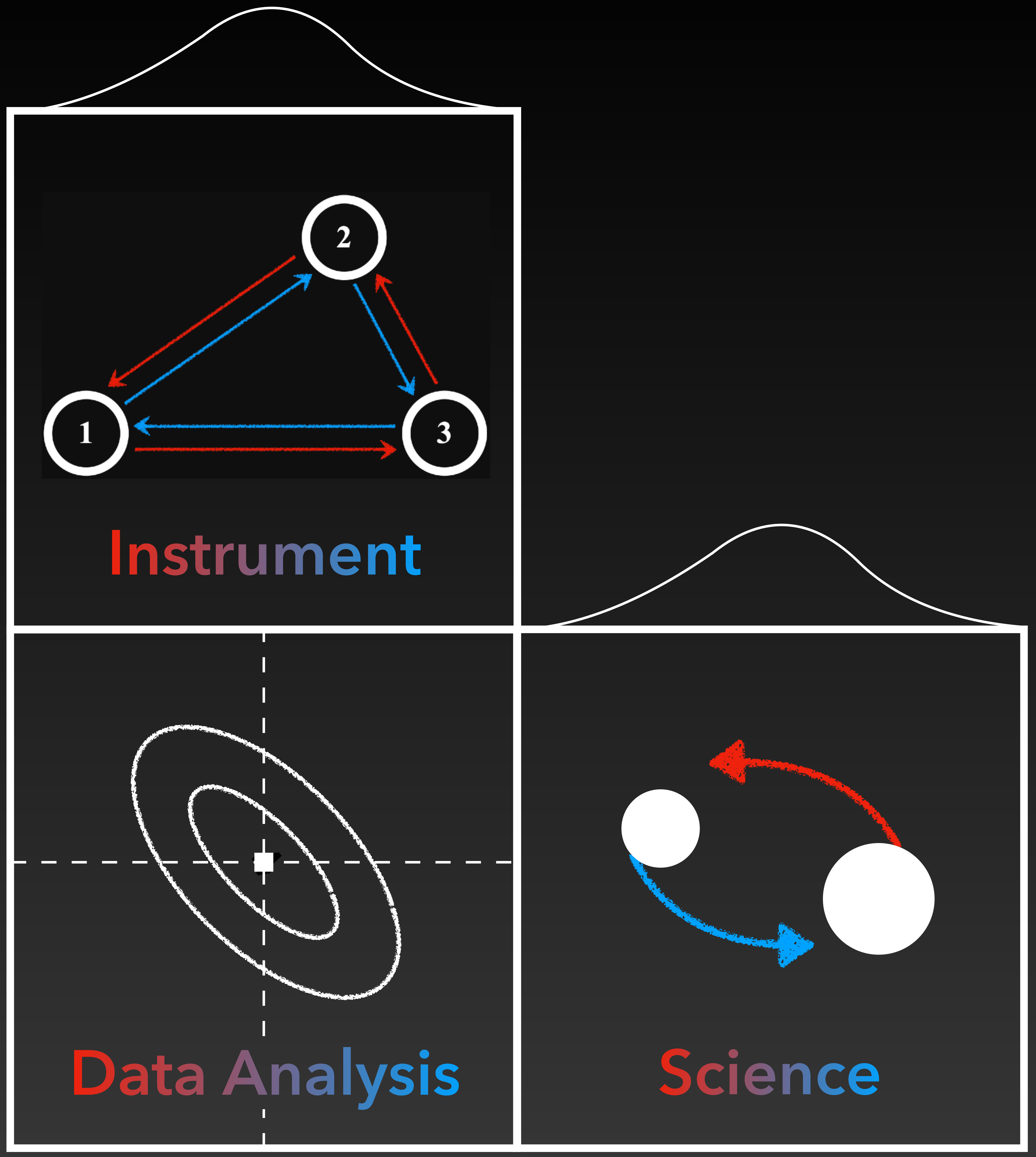


Taiji Data Challenge

Exploring & addressing the new challenges for space-based GW data analysis

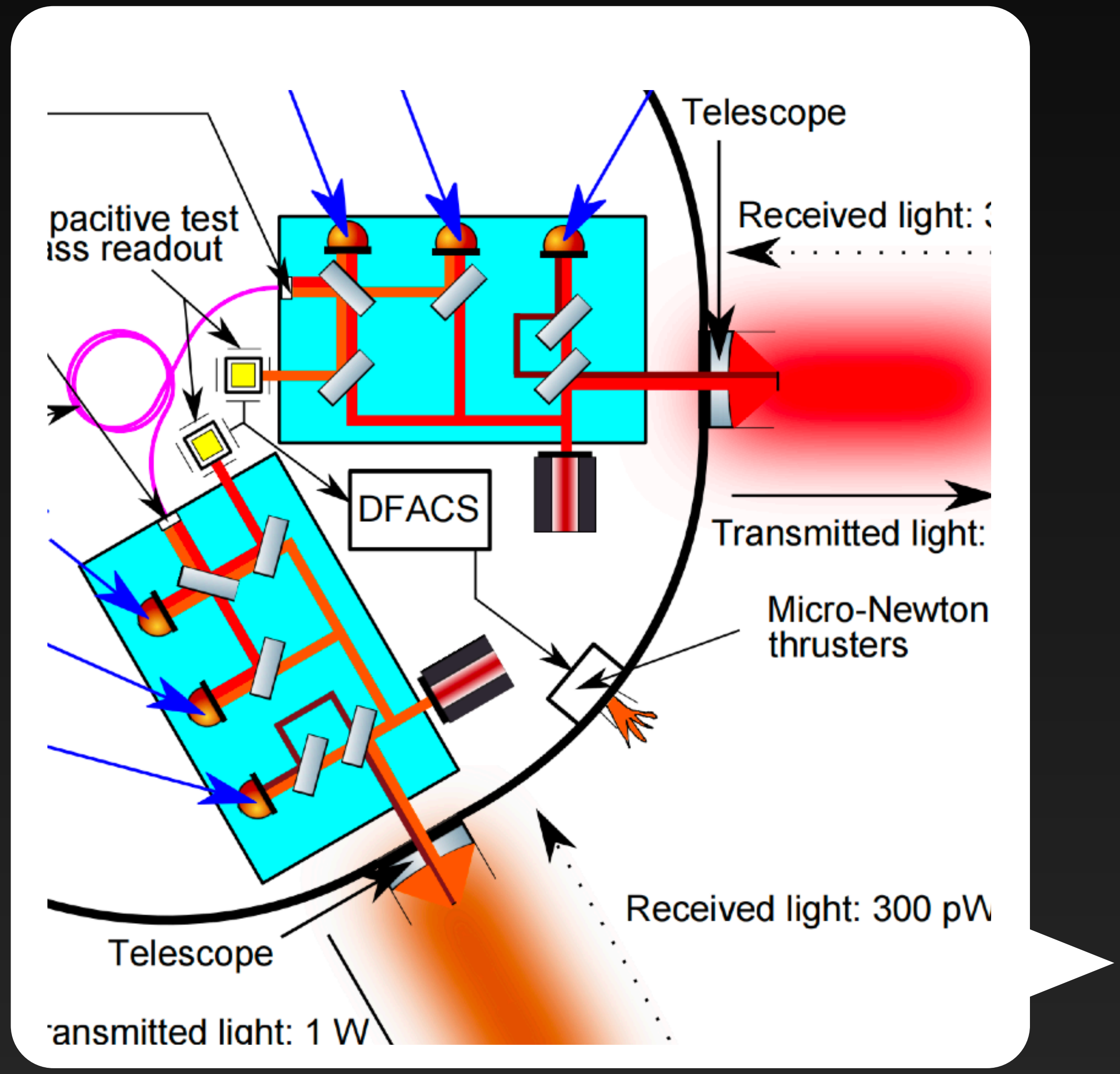
The TDC II working group
UCAS, IMECH CAS, MICROSATE CAS, SHAO CAS, NSSC CAS, PKU, BNU

1. Basic Concepts
2. Mock Data Pipeline
3. Challenges
4. The TDC Datasets
5. Todos
6. Triangle
7. Tests and Verifications
8. Take-away messages

1. **Basic Concepts**
2. Mock Data Pipeline
3. Challenges
4. The TDC Datasets
5. Todos
6. Triangle
7. Tests and Verifications
8. Take-away messages

Basic Concepts: the Taiji Detector

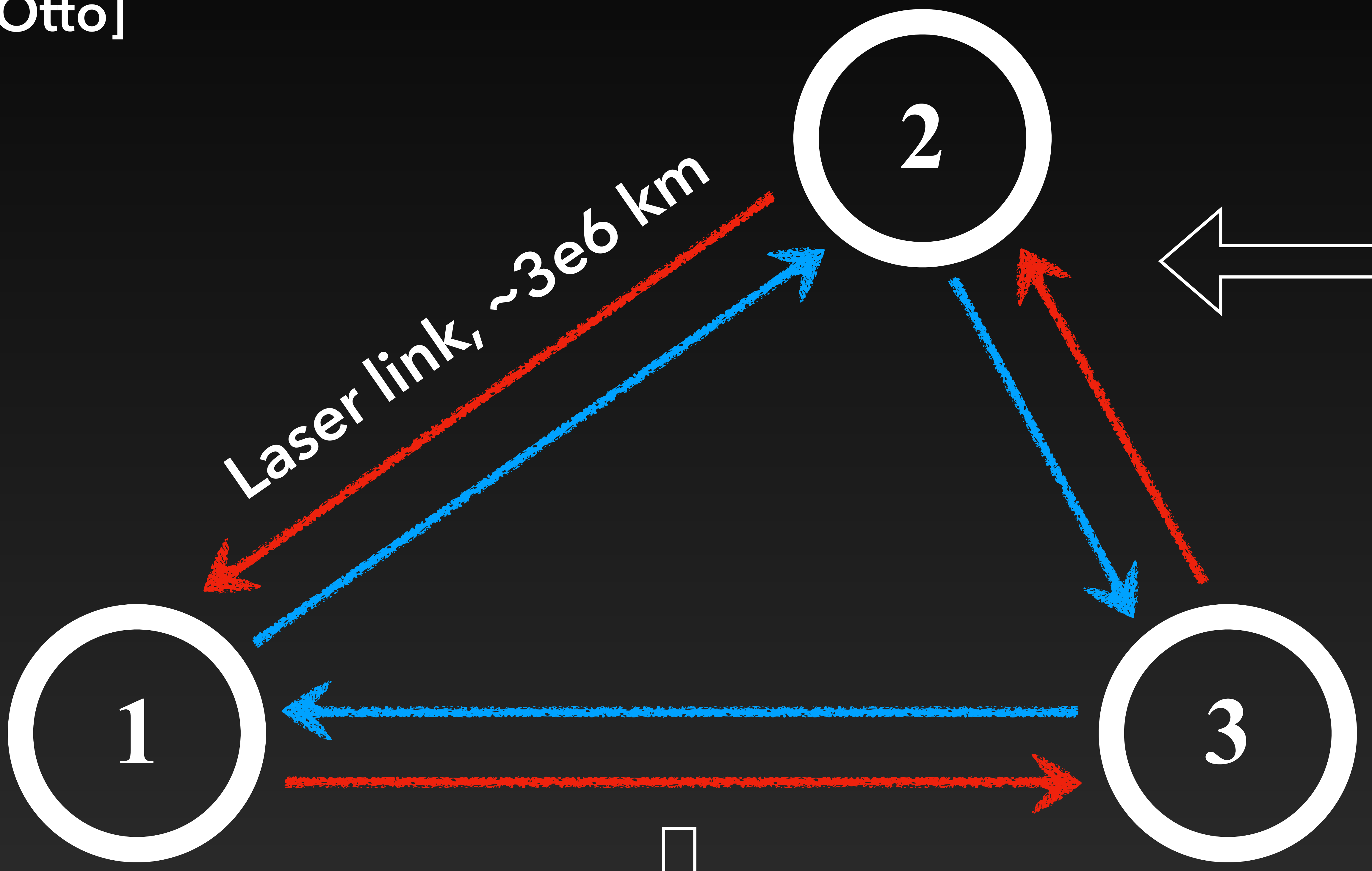
[Image Credit: PhD thesis of M. Otto]

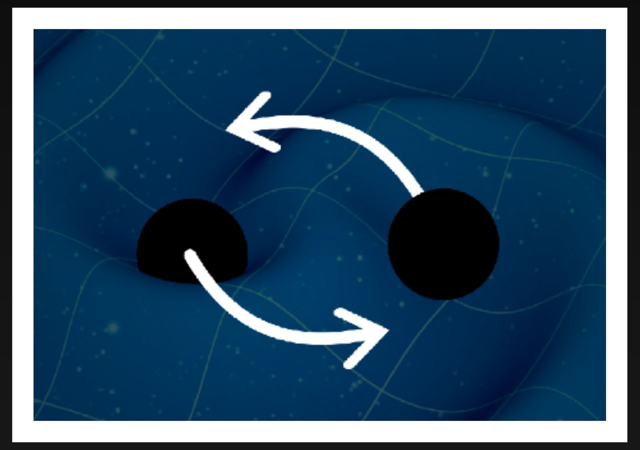


Inside the SCs

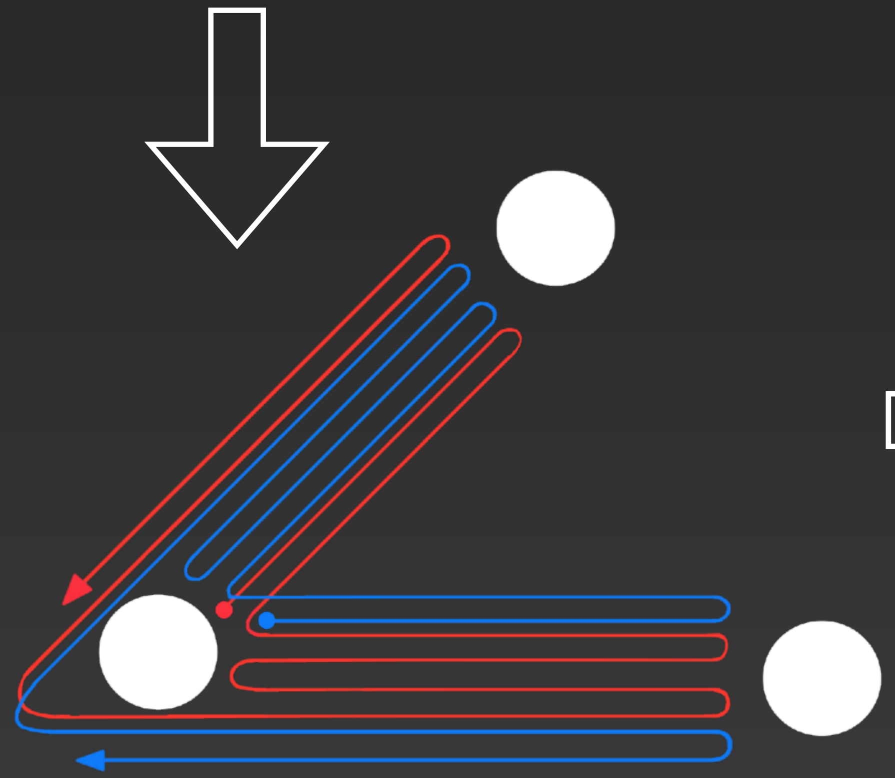
- Telescopes to receive and send lasers
- Test-masses (TMs) as inertial references
- Interferometers to readout the optical path changes
- Ultra-stable laser sources
- … …

Spacecraft (SC)

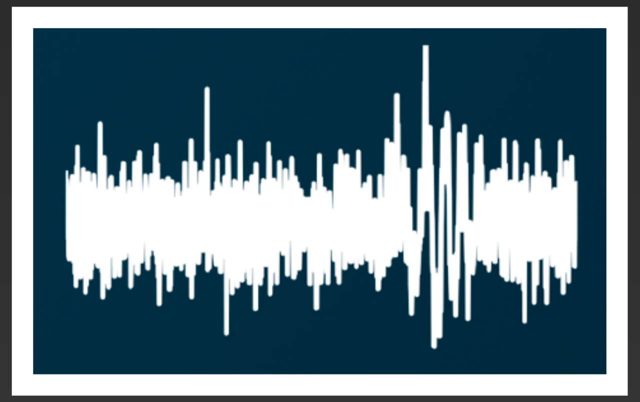




GW source



Time-delay interferometry



TDI data

Basic Concepts: the Targets of Taiji

- Massive black hole binaries (MBHBs)

  $m \in (10^4, 10^7) M_\odot$, $\mathcal{O}(10) \sim \mathcal{O}(100)$ per year

- Galactic binaries (GBs)

  $m < 1.4 M_\odot$ (for DWDs), $\mathcal{O}(10^7)$

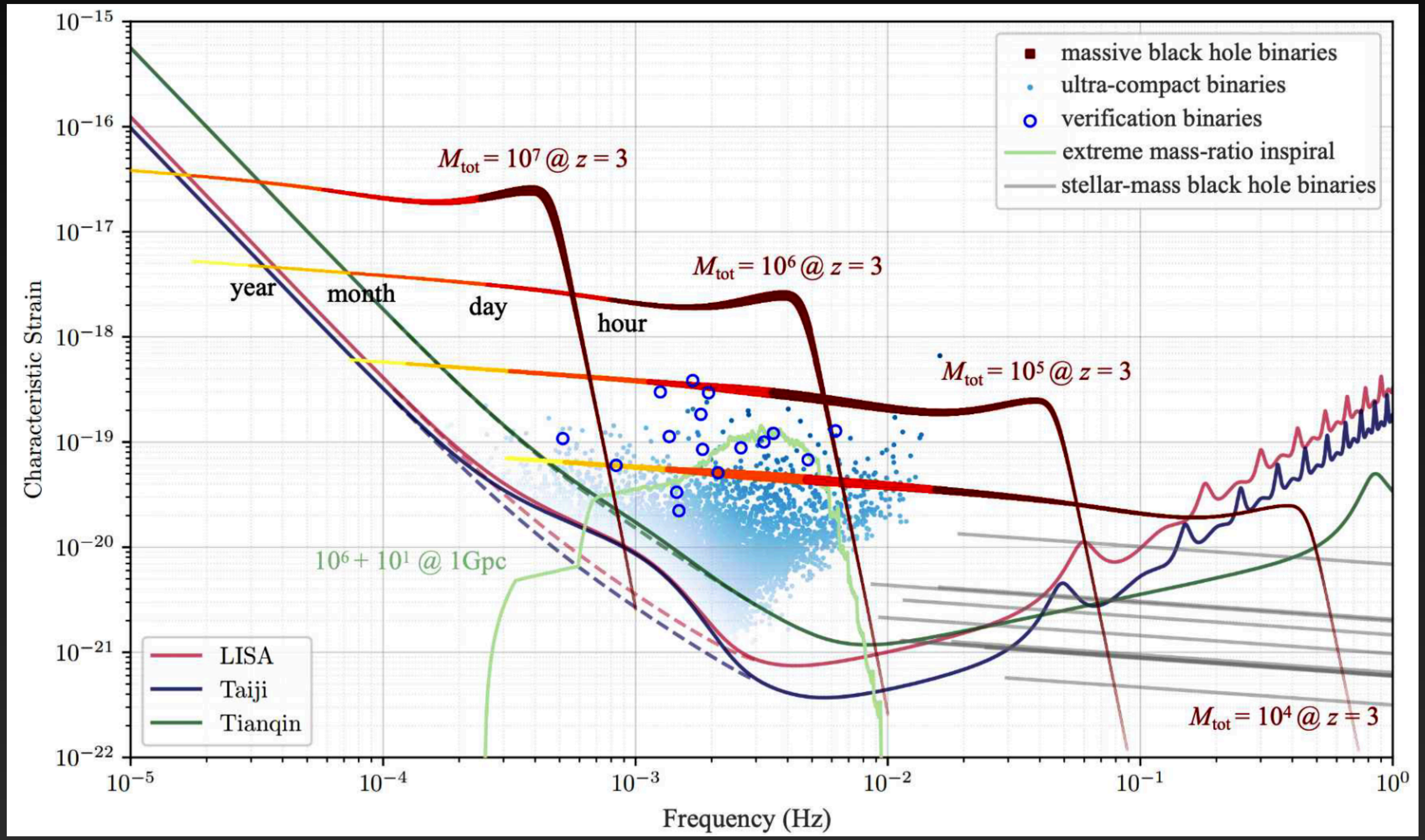
- Extreme mass-ratio inspirals (EMRIs)

  $m_2/m_1 \in (10^{-3}, 10^{-6})$, $\mathcal{O}(1) \sim \mathcal{O}(10^3)$

- Stellar mass black hole binaries (SMBHBs)

  $m \in (5, 80) M_\odot$

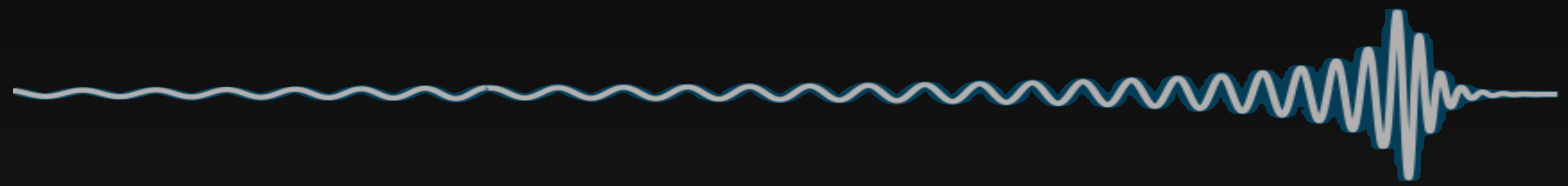
- Stochastic GW background (SGWB)
- …



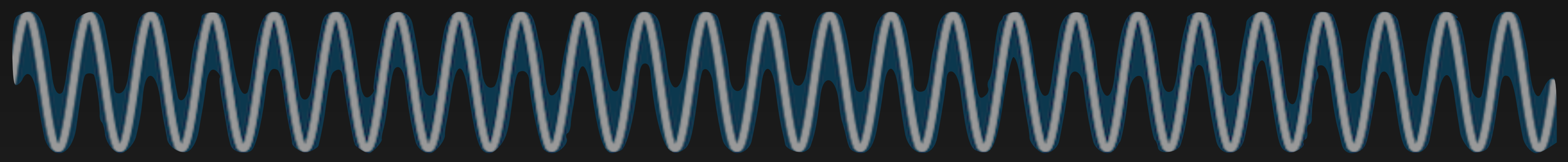
Target sources and sensitivities of LISA, Taiji, Tianqin

Basic Concepts: the Targets of Taiji

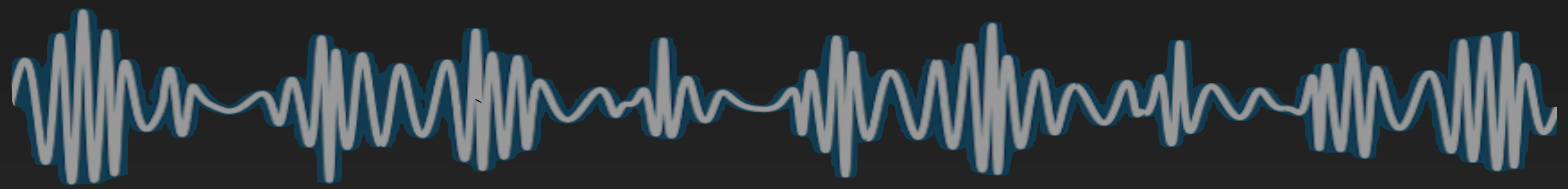
- Massive black hole binaries (MBHBs)



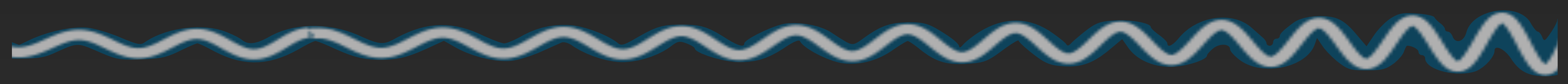
- Galactic binaries (GBs)



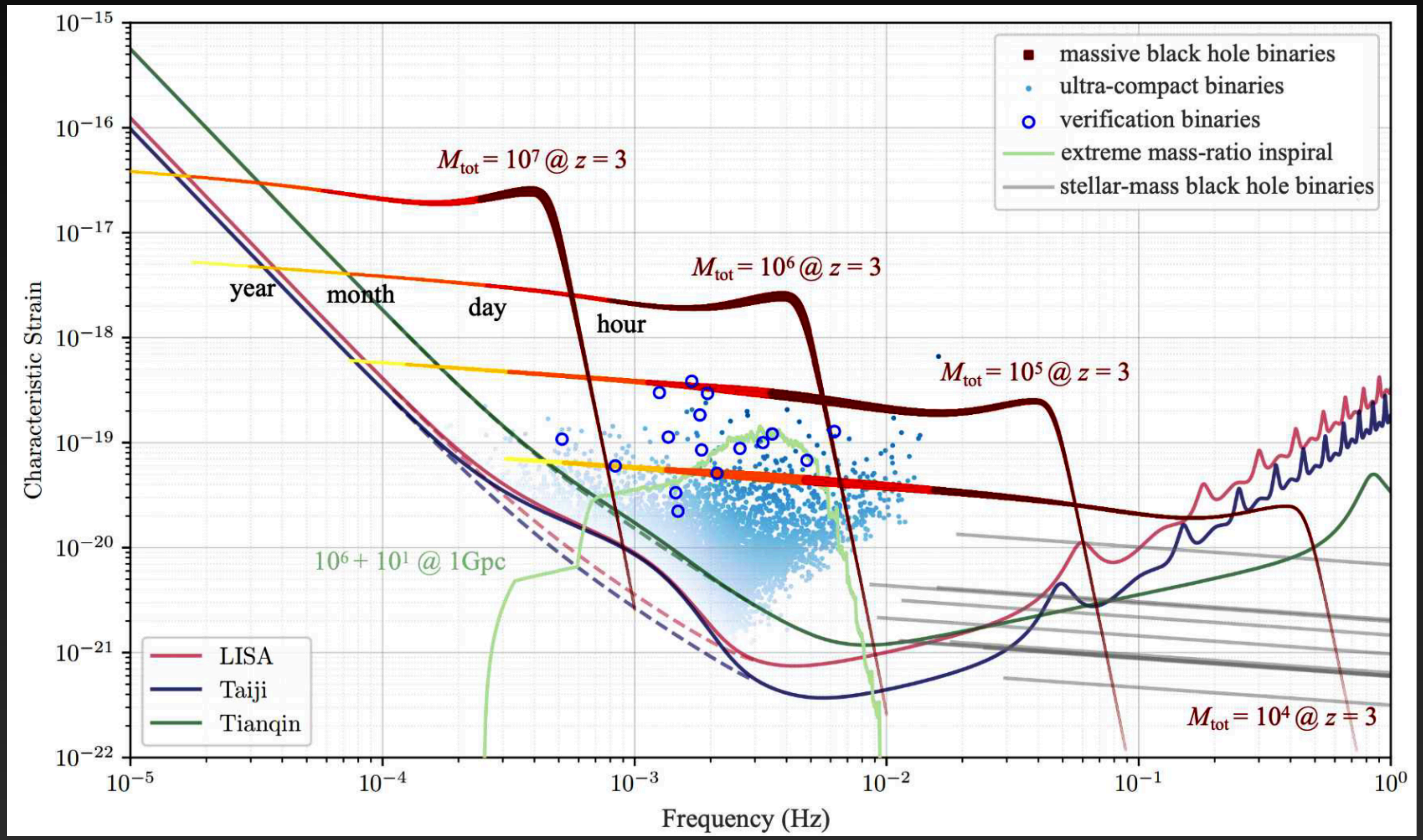
- Extreme mass-ratio inspirals (EMRIs)



- Stellar mass black hole binaries (SMBHBs)

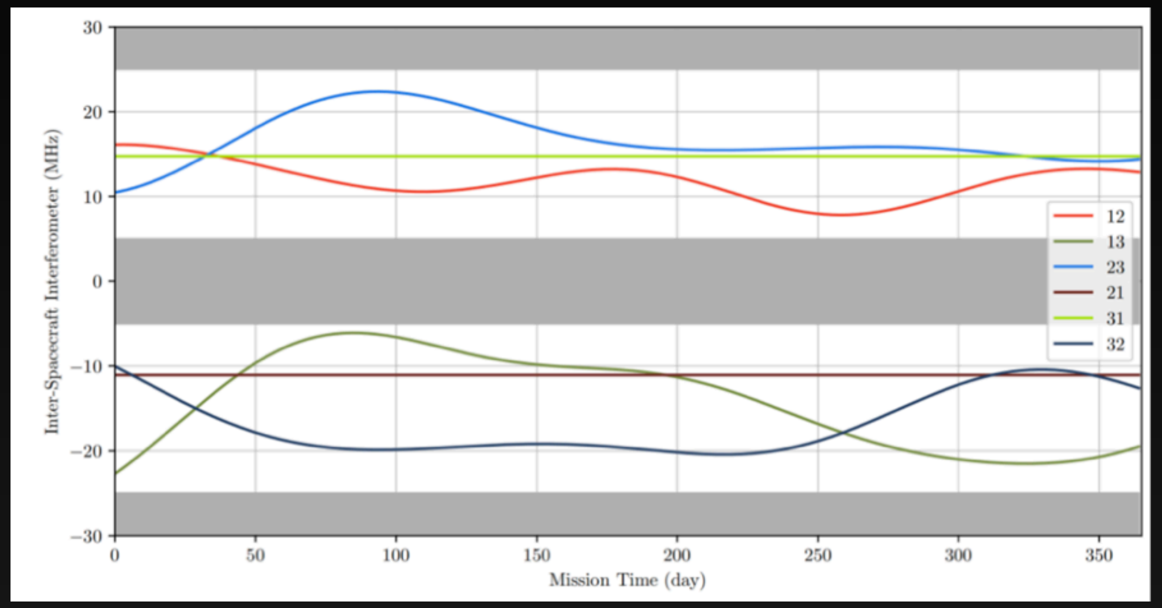


- Stochastic GW background (SGWB)
- …

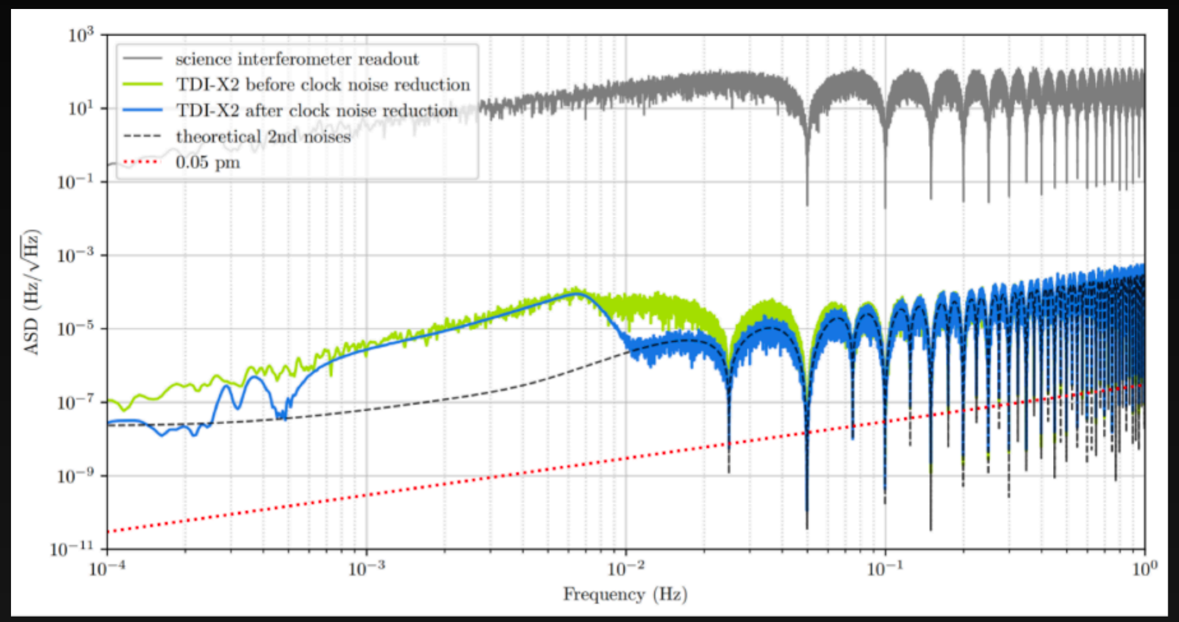


Target sources and sensitivities of LISA, Taiji, Tianqin

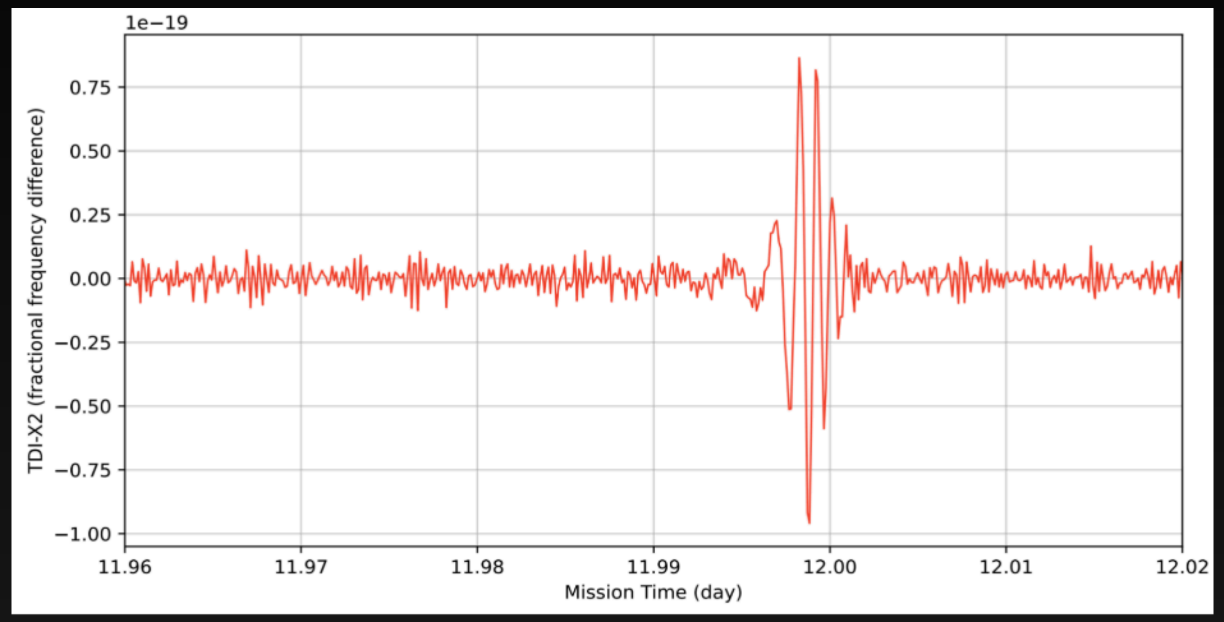
Basic Concepts: Taiji data flow



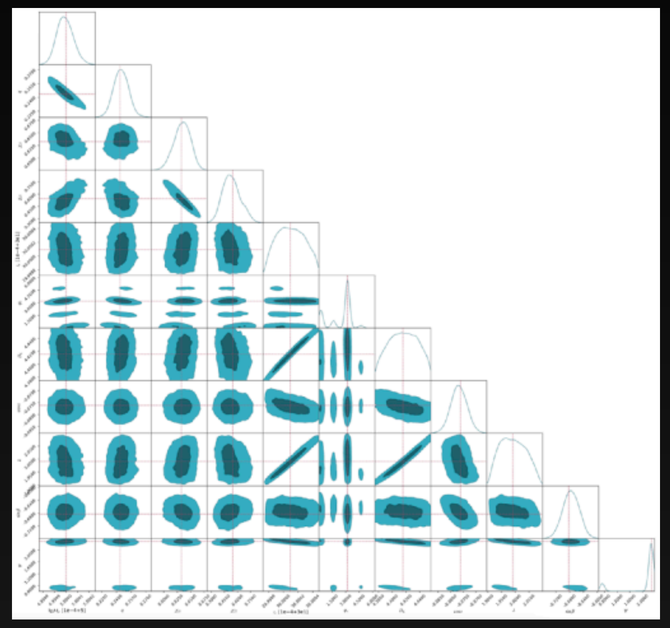
L0: Raw interferometric data



TDI processing



L1: Science data



L2,3: GW science

L0 data → **Pre-processing** → L1 data → **Scientific analysis** → L2,3 data

- **Inter-spacecraft ranging, clock synchronization**
- **Noise suppression**

Laser noises (8 orders larger than GWs), clock noises (6 orders larger than GWs), SC jitters, tilt-to-length (TTL) coupling noises …

- **Calibration of key operation parameters**

Gravitational reference sensor (GRS) center-of-mass offset, GRS scale factor, GRS acceleration bias, …

- **Assessment of operational status and data quality**

- **Low latency alert pipeline**

Merger time, sky locations, early-warning for follow-up GW and electro-magnetic observations

- **Global fit analysis pipeline**

10^4 detectable GW sources, 10^5 GW source parameters, O(?) instrumental & astrophysical noise parameters

- **Scientific interpretations**

Source population model, Galaxy evolution, cosmic history …

Basic Concepts: data challenge

The aims of data challenges:

- Discover and address the challenges in data analysis in advance
- Develop and test algorithms and tools for Taiji's scientific application system
- Provide tools for GW sciences

Data challenges in the past

LISA Data Challenge: https://lisa-ldc.lal.in2p3.fr/

Taiji Data Challenge I: http://taiji-tdc.ictp-ap.org/

GWSpace: https://github.com/TianQinSYSU/GWSpace

Simplified Orbit

(equal-arm analytic orbit, Kepler orbit)

Simplified Noises

(2 or 3 noise components, Gaussian stationary)

Simplified Signals

(simple waveforms and responses)

…… **Mildly overlapping**

~~Analysis tools~~

~~for realistic data~~

What's in Taiji Data Challenge II

- **Data**: 5 groups of datasets
- **Codes**:

Triangle-Simulator: time-domain simulation, costly but accurate, benchmark

Triangle-GB: fast frequency-domain GB analysis, illustrative example

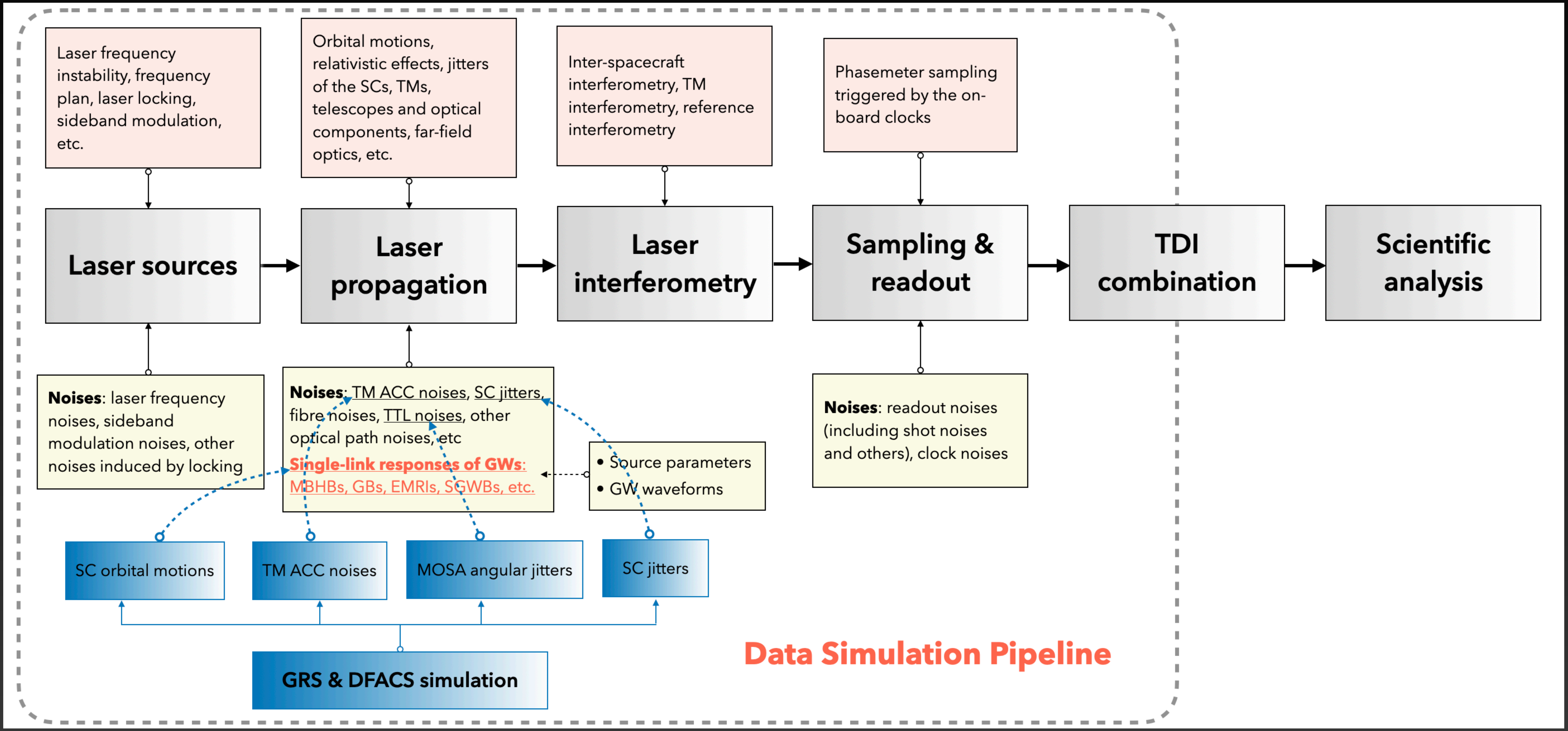
Triangle-BBH: fast frequency-domain MBHB analysis (PhenomD/HM), illustrative example

Break down to tasks

- Signal response & noise modeling & programming
- Statistical analysis
- Bayesian inference & stochastic sampling
- … ….

1. Basic Concepts
2. **Mock Data Pipeline**
3. Challenges
4. The TDC Datasets
5. Todos
6. Triangle
7. Tests and Verifications
8. Take-away messages

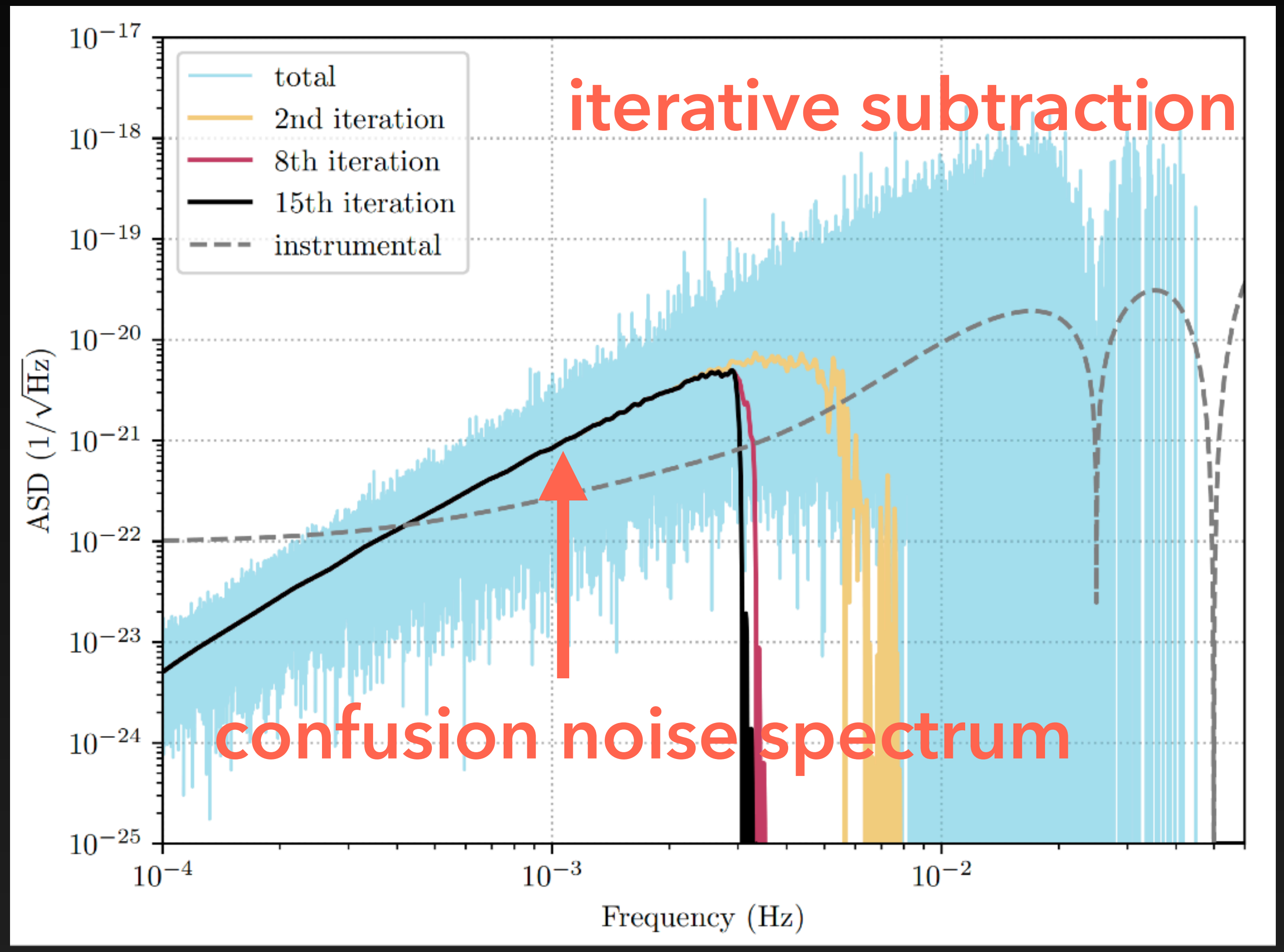
Mock Data Pipeline

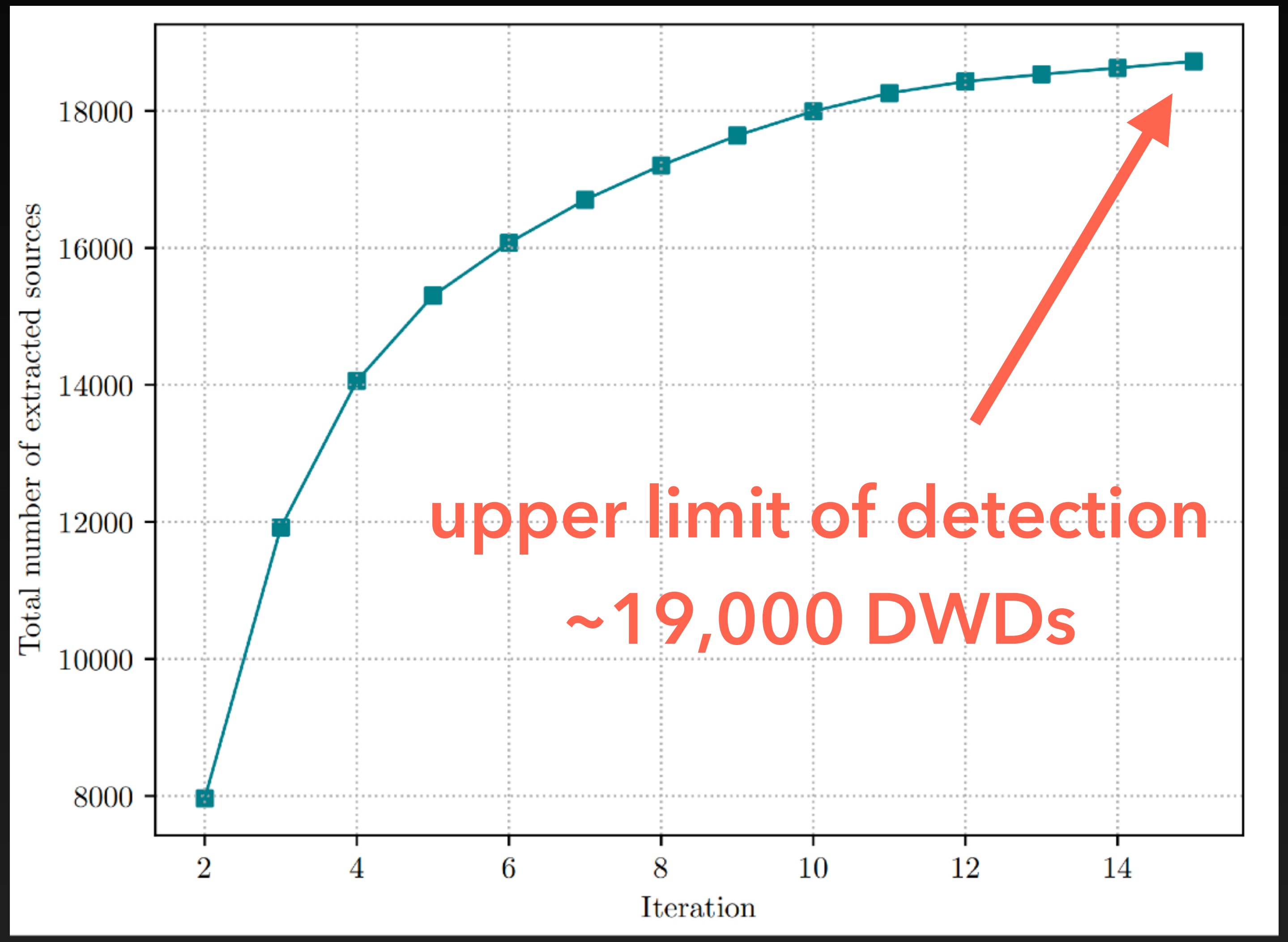


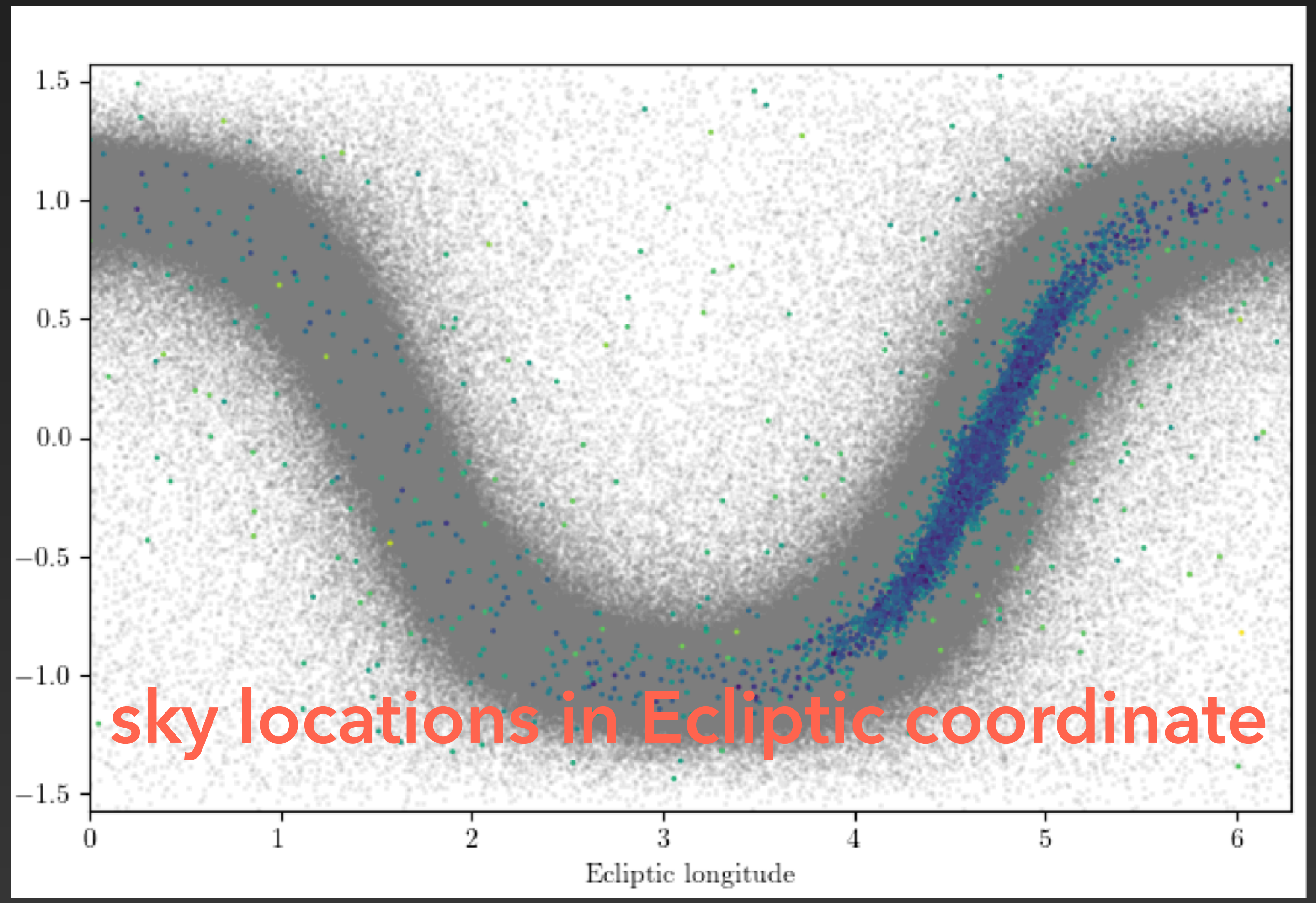
Models Behind the TDC Data

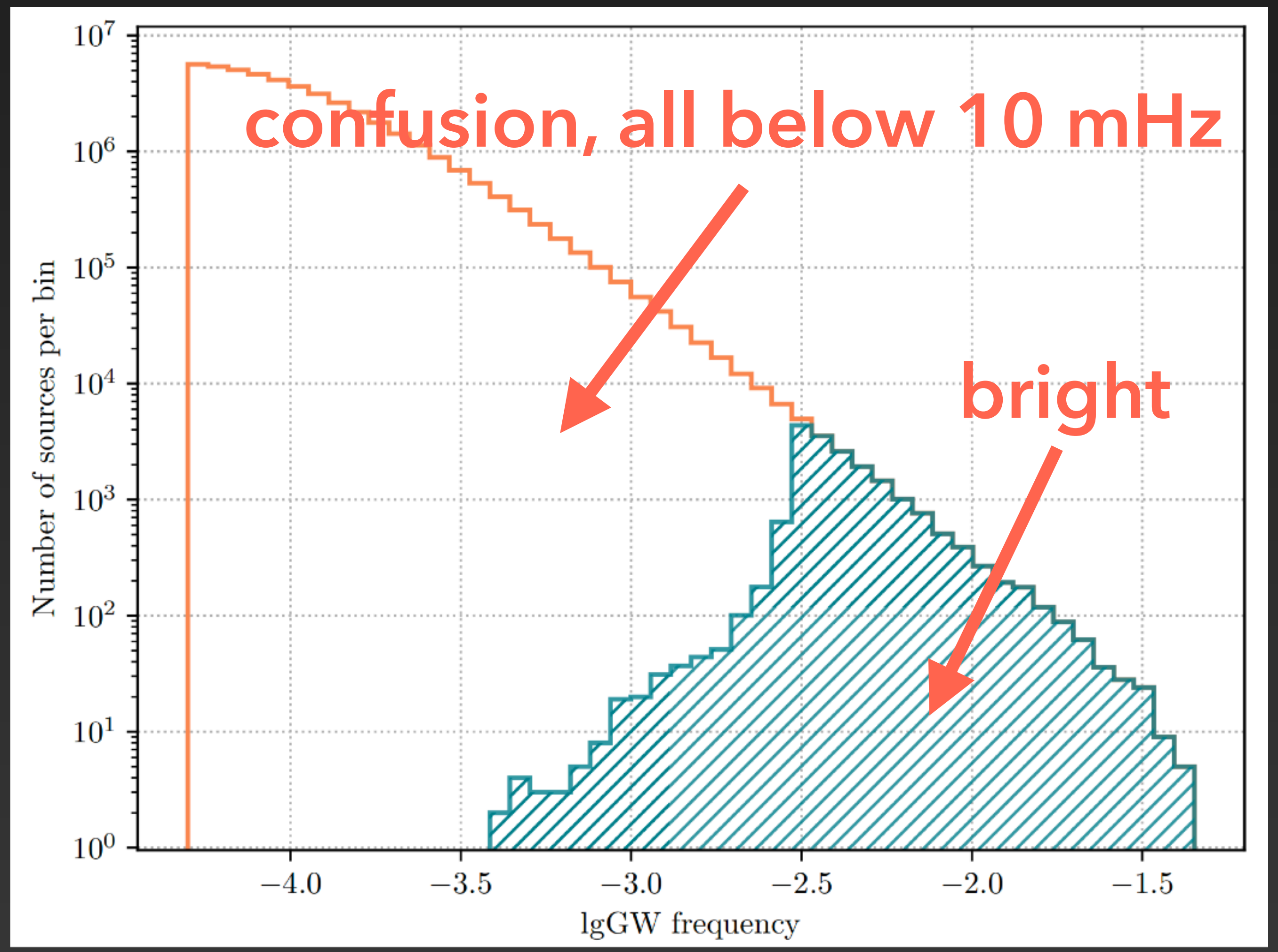
The TDC Populations

- **MBHB**: mix of 3 populations (PopIII, Q3delay, Q3nodelay), normalized according to Ref. [1]
- **GB**:
  - 55 LISA verification binaries [2]
  - observation-driven DWD population model of Ref. [3] adjusted for Taiji (lower limit 0.1 mHz -> 0.05 mHz), ~45 million in total
  - Bright & confusion GB catalogue generated via iterative subtraction [4]
- **EMRI**: typical source parameters









Models Behind the TDC Data

The TDC Waveforms

- **MBHB**: IMRPhenomD, IMRPhenomT, SEOBNRv5 -EHM [5]
- **GB**: Taylor expansion to the second derivative of frequencies
- **EMRI**: AK, Kerr-BH,Bumpy-BH, b-EMRI waveforms [6] by SHAO
- **SGWB**: power-law for astrophysical SGWB, double broken power-law for cosmological SGWB (FoPT) [7]
- **Detector response**: time-domain response based on numerical orbit, no equal-arm or low-frequency or static constellation approximation

Template = Waveform + Projection + TDI

where orbit information enters

$$\eta_{ij}(t) \equiv \frac{\nu_{\text{receive}} - \nu_{\text{send}}}{\nu_{\text{send}}} \approx \frac{1}{2\left[1 - \hat{\boldsymbol{k}} \cdot \boxed{\hat{\boldsymbol{n}}_{ij}(t)}\right]} \left[\Phi_{ij}\left(t - \frac{\boxed{d_{ij}(t)}}{c} - \frac{\hat{\boldsymbol{k}} \cdot \boxed{\boldsymbol{R}_j(t)}}{c}\right) - \Phi_{ij}\left(t - \frac{\hat{\boldsymbol{k}} \cdot \boxed{\boldsymbol{R}_i(t)}}{c}\right)\right], \quad \Phi_{ij}(t) = \boxed{\hat{\boldsymbol{n}}_{ij}(t) \otimes \hat{\boldsymbol{n}}_{ij}(t)} : \boldsymbol{h}(t)$$

$$X_2 = \left(1 - \boxed{\mathbf{D}_{121} - \mathbf{D}_{12131} + \mathbf{D}_{1312121}}\right)\left(\eta_{13} + \boxed{\mathbf{D}_{13}}\eta_{31}\right) - \left(1 - \boxed{\mathbf{D}_{131} - \mathbf{D}_{13121} + \mathbf{D}_{1213131}}\right)\left(\eta_{12} + \boxed{\mathbf{D}_{12}}\eta_{21}\right)$$

Models Behind the TDC Data

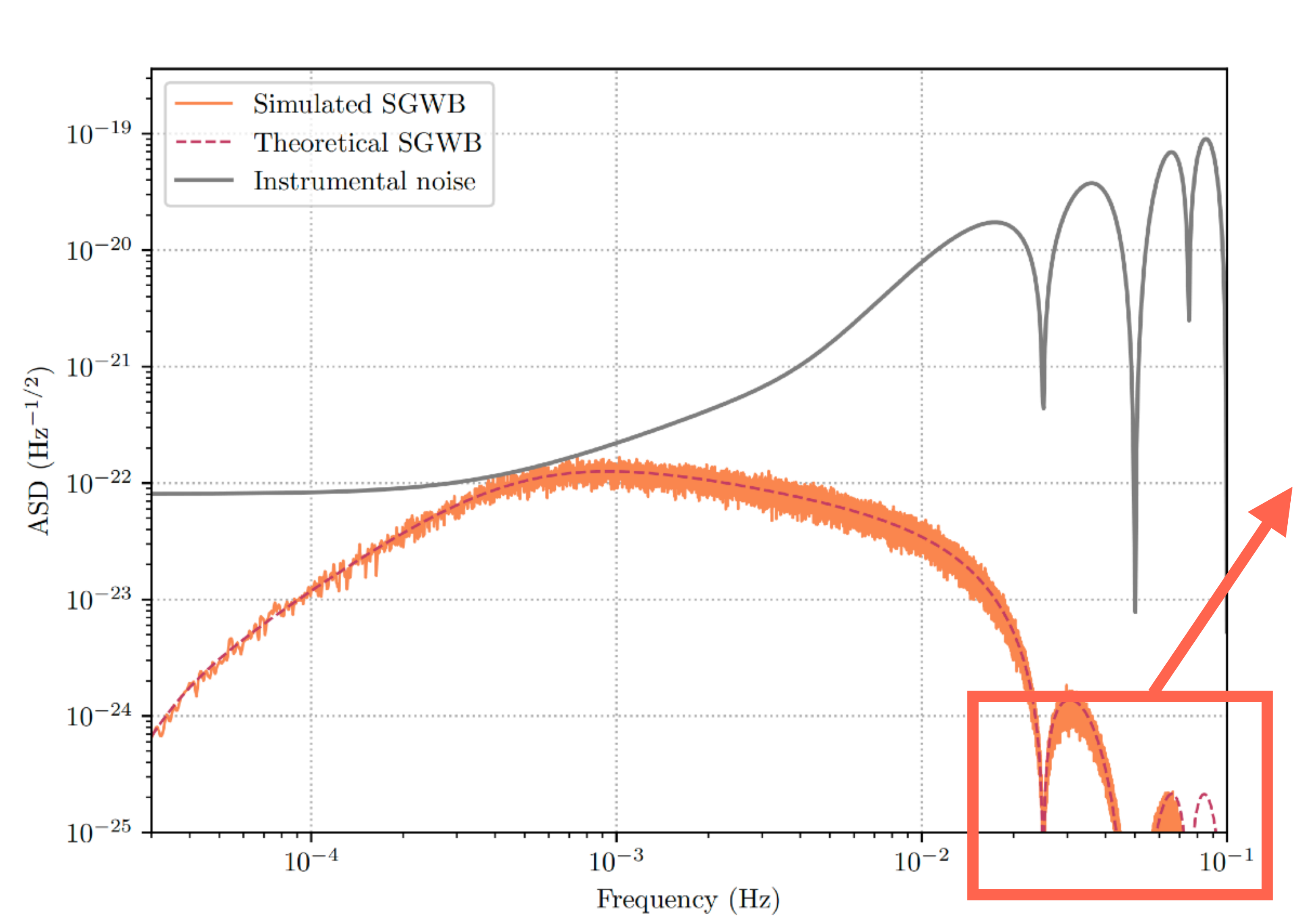
The TDC Waveforms

- **MBHB**: IMRPhenomD, IMRPhenomT, SEOBNRv5 -EHM [5]
- **GB**: Taylor expansion to the second derivative of frequencies
- **EMRI**: AK, Kerr-BH,Bumpy-BH, b-EMRI waveforms [6] by SHAO
- **SGWB**: power-law for astrophysical SGWB, double broken power-law for cosmological SGWB (FoPT) [7]
- **Detector response**: time-domain response based on numerical orbit, no equal-arm or low-frequency or static constellation approximation

Calculation of SGWB responses:

- Divide the whole celestial sphere with **Healpix**
- Generate stochastic signals at each direction based on Ω_{GW}
- Response in time domain

Slow, but can account for time-varying response & anisotropy SGWBs

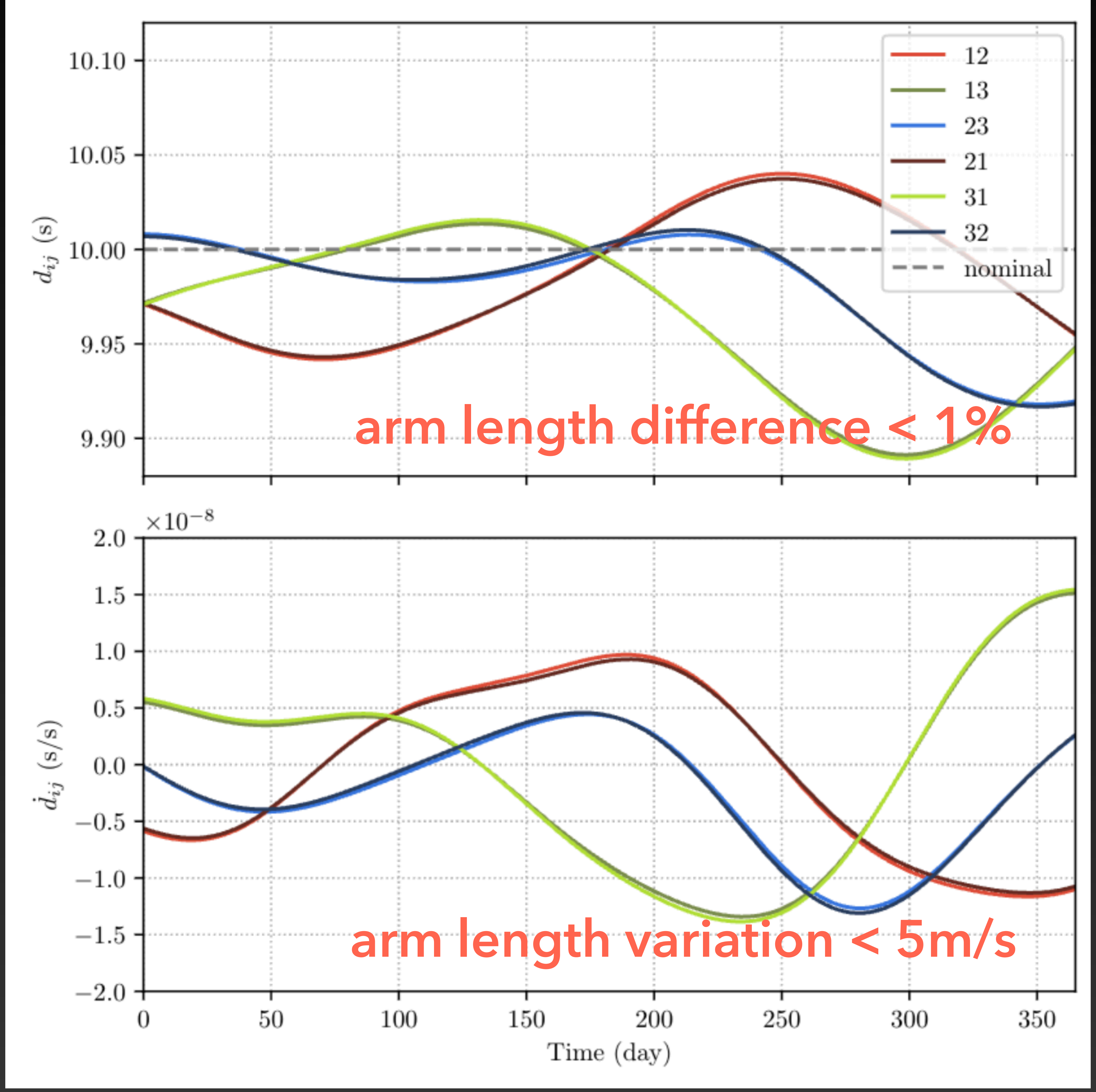


Not perfectly matched since the theory is based on low-frequency equal-arm approximation

Models Behind the TDC Data

The TDC Orbit

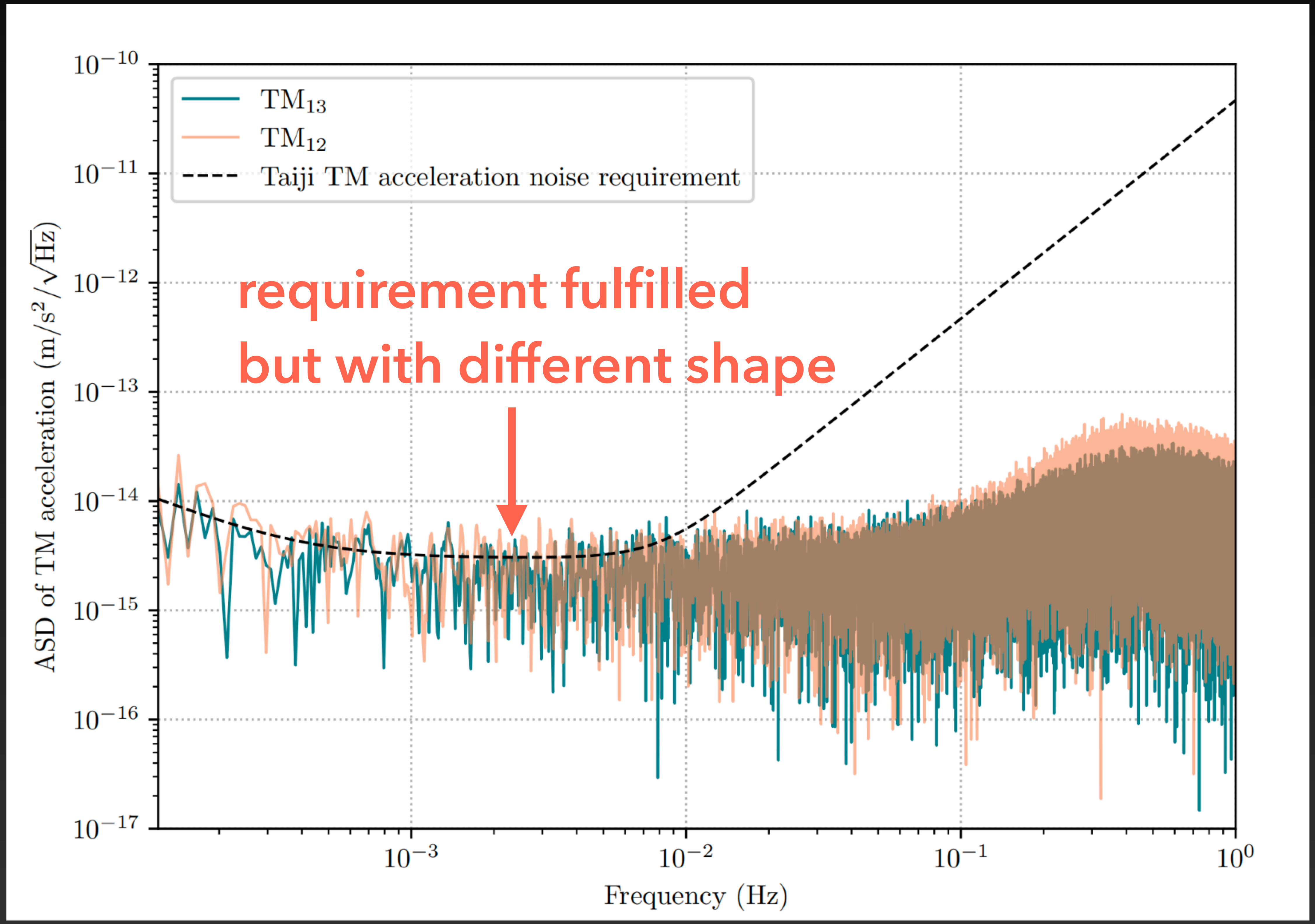
- The numerical orbit is based on the drag-free attitude control system (DFACS) simulation of MICROSATE, CAS
- **Perturbations**: solar pressure, celestial gravity, micro-thrusters, IFO & DWS sensing noises, etc.
- **Length**: 1 year



Models Behind the TDC Data

The TDC Noises

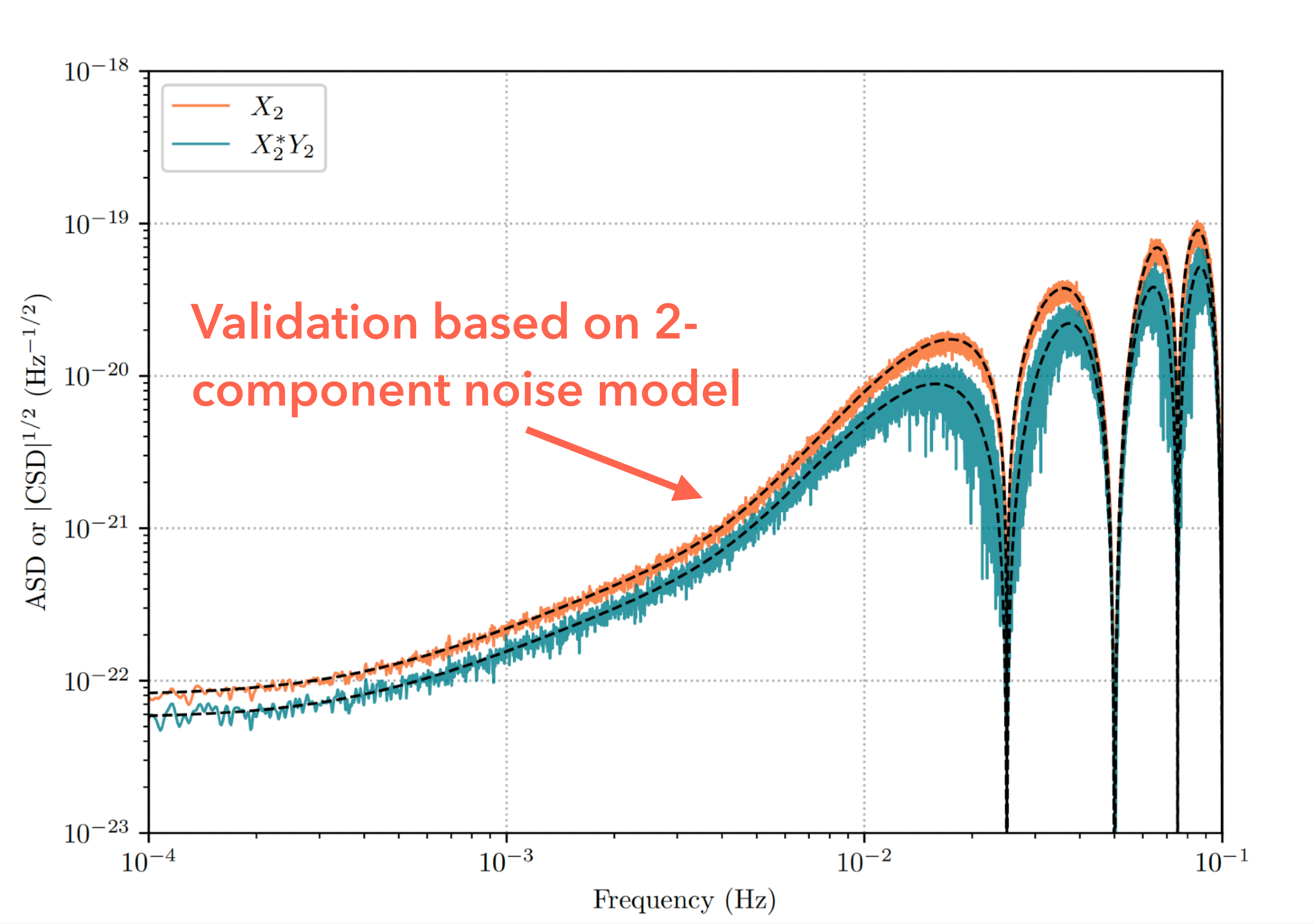
- Noises generated from **analytical models**: laser noises, clock noises, IFO readout noises, fibre noises, other optical path noises, sideband modulation noises, pseudo ranging noises
- Noises extracted from the **DFACS simulation**: TM acceleration noises, SC jitters, angular jitters of SC and MOSAs
- glitches and gaps can be customized



Models Behind the TDC Data

The TDC Noises

- Noises generated from **analytical models**: laser noises, clock noises, IFO readout noises, fibre noises, other optical path noises, sideband modulation noises, pseudo ranging noises
- Noises extracted from the **DFACS simulation**: TM acceleration noises, SC jitters, angular jitters of SC and MOSAs
- glitches and gaps can be customized



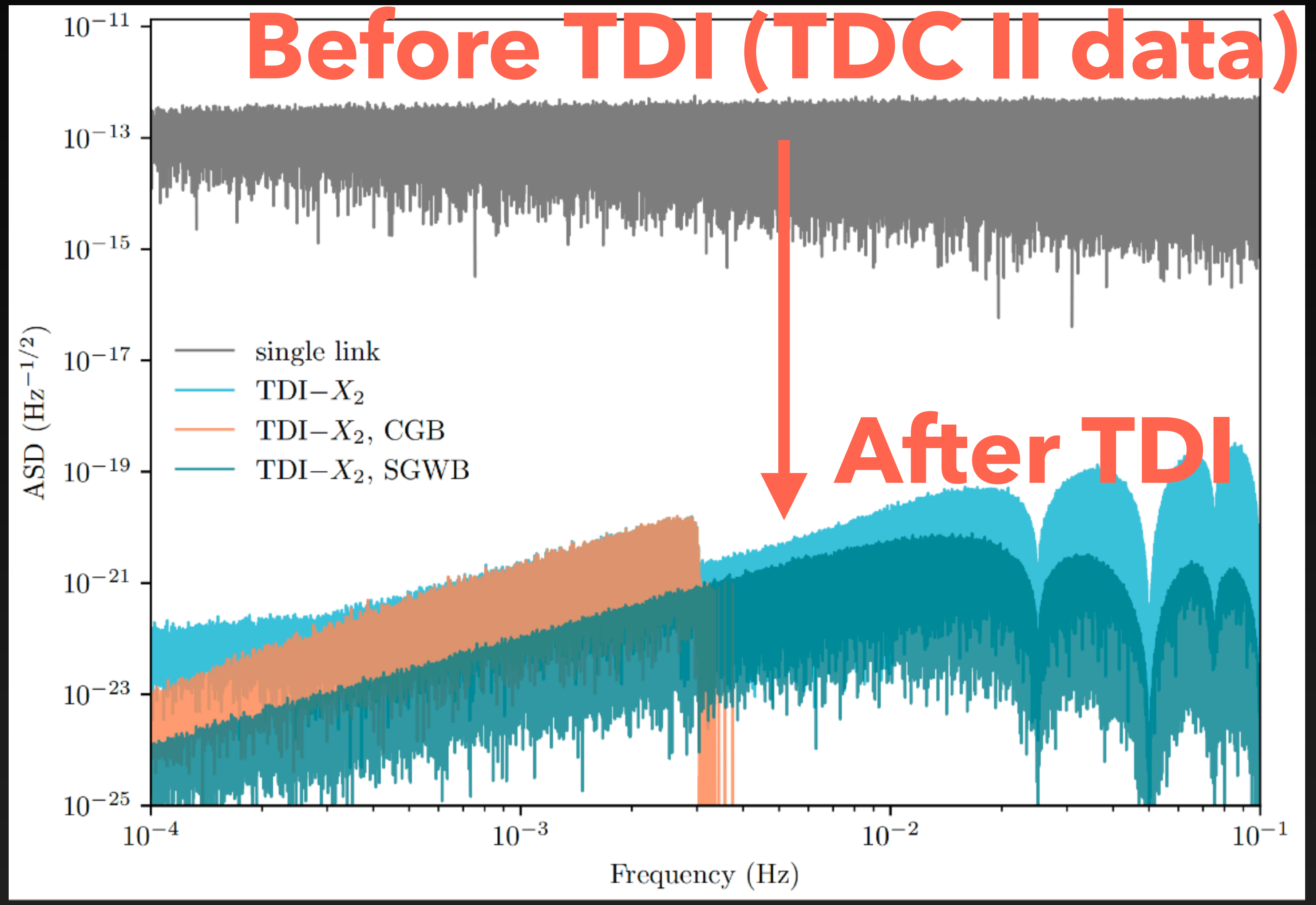
Models Behind the TDC Data

The TDC Data

- **Data length**: 1 year (for most datasets), 90 days, 30 days
- **Sampling rate**: 1 Hz (for most datasets), 4 Hz
- **Time frame**: Barycentric coordinate time (for most datasets), on-board clock times
- **Data form**: single-link interferometry eta (for most datasets), raw interferometric measurements + auxiliary data

$$\mathrm{TDI} = \sum_{ij \in \mathrm{MOSAs}} \mathbf{P}_{ij}\eta_{ij}$$

$$\begin{aligned}
&\mathbf{P}_{12} = 1 - \mathbf{D}_{131} - \mathbf{D}_{13121} + \mathbf{D}_{1213131},\\
&\mathbf{P}_{23} = 0,\\
&\mathbf{P}_{31} = -\mathbf{D}_{13} + \mathbf{D}_{1213} + \mathbf{D}_{121313} - \mathbf{D}_{13121213},\\
&\mathbf{P}_{21} = \mathbf{D}_{12} - \mathbf{D}_{1312} - \mathbf{D}_{131212} + \mathbf{D}_{12131312},\\
&\mathbf{P}_{32} = 0,\\
&\mathbf{P}_{13} = -1 + \mathbf{D}_{121} + \mathbf{D}_{12131} - \mathbf{D}_{1312121}.
\end{aligned}$$



References:

[1] A. Mangiagli et al, Phys. Rev. D 106, 103017, 2022

[2] T. Kupfer et al, arXiv:2302.12719, https://gitlab.in2p3.fr/LISA/lisa-verification-binaries

[3] V. Korol et al, Mon.Not.Roy.Astron.Soc. 511 (2022) 4, 5936-5947

[4] N. Karnesis et al, Phys. Rev. D 104, 043019, 2021

[5] A. Gamboa et al, arXiv:2412.12831

[6] P. Shen et al, Phys. Rev. D 108, 064015, 2023, Phys. Rev. D 111, 024004, 2025

[7] R. Rosati et al, arXiv:2410.17180

1. Basic Concepts
2. Mock Data Pipeline
3. **Challenges**
4. The TDC Datasets
5. Todos
6. Triangle
7. Tests and Verifications
8. Take-away messages

Challenges

1. The complexity of signal modeling

Matched filter

$$(d\,|\,h) \equiv 4\mathrm{Re}\int_{f_{\min}}^{f_{\max}} \frac{\tilde{d}(f)\tilde{h}^*(f)}{S(f)}\mathrm{d}f$$

data

template

noise

Template = Waveform + Projection + TDI

What makes an "accurate" template? Criteria in the literature:

$$(h_1 - h_0\,|\,h_1 - h_0) < 1 \Leftrightarrow 1 - \mathcal{O}(h_0, h_1) < \frac{D}{2\rho_0^2} \Leftrightarrow \Delta\theta_a^{\rm sys} = -\sum_{\theta_b}\left(\Gamma^{-1}\right)_{\theta_a\theta_b}\left(\frac{\partial h_1}{\partial\theta_b}\middle|\, h_1 - h_0\right) < \Delta\theta_a^{\rm stat} = \sqrt{\left(\Gamma^{-1}\right)_{\theta_a\theta_a}}$$

It is more difficult to make unbiased estimations for high-SNR signals. $\Delta\theta_a^{\rm stat}\downarrow \Rightarrow \Delta\theta_a^{\rm sys}/\Delta\theta_a^{\rm stat}\uparrow$

Publications on this topic:

[1] M. Pürrer et al, Phys. Rev. Research, 2020

[2] L. Lindblom et al, Phys. Rev. D 78, 2008

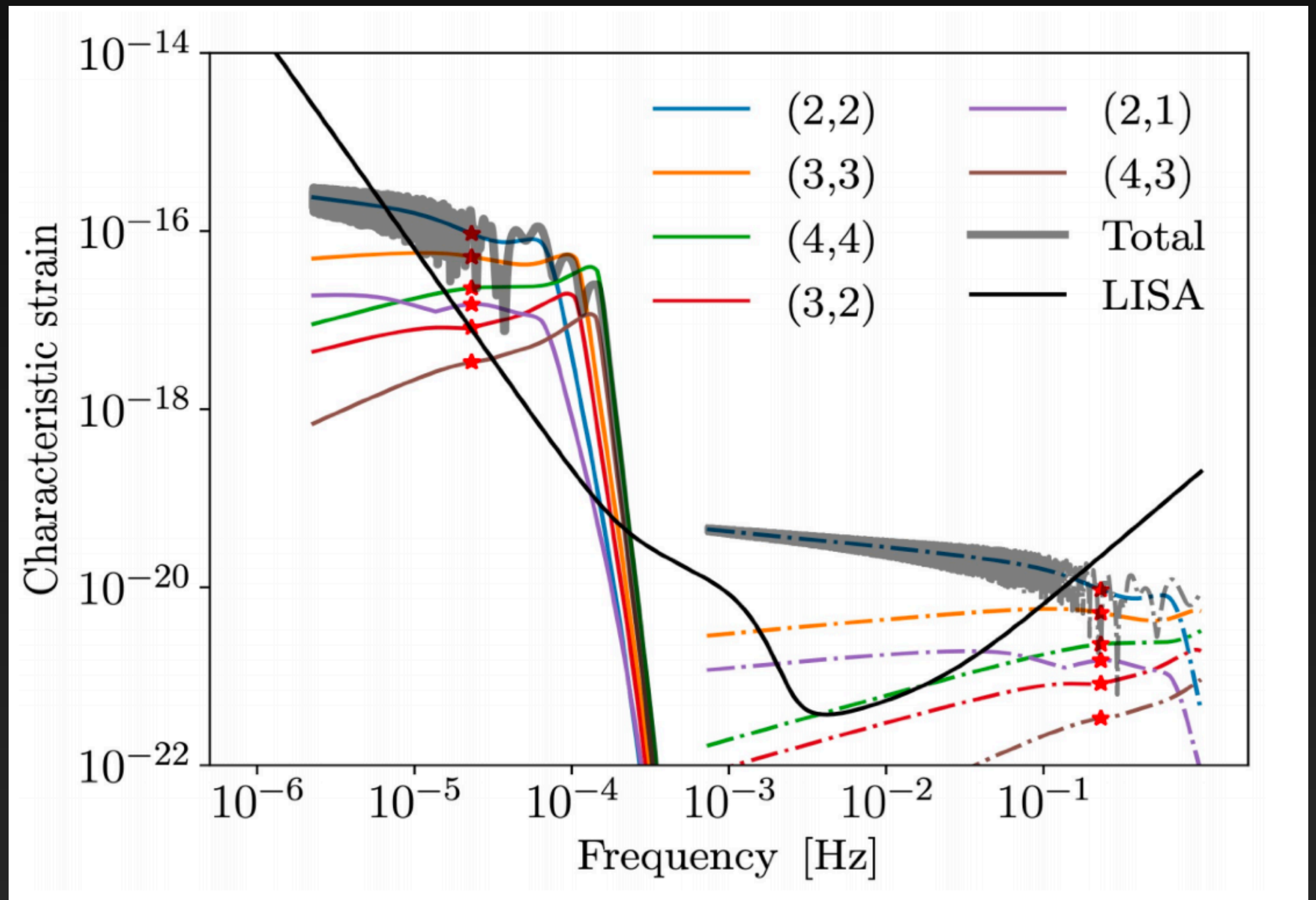
[3] K. Chatziioannou et al, Phys. Rev. D 95, 104004, 2017

[4] V. Kapil et al, Phys. Rev. D 109, 104043, 2024

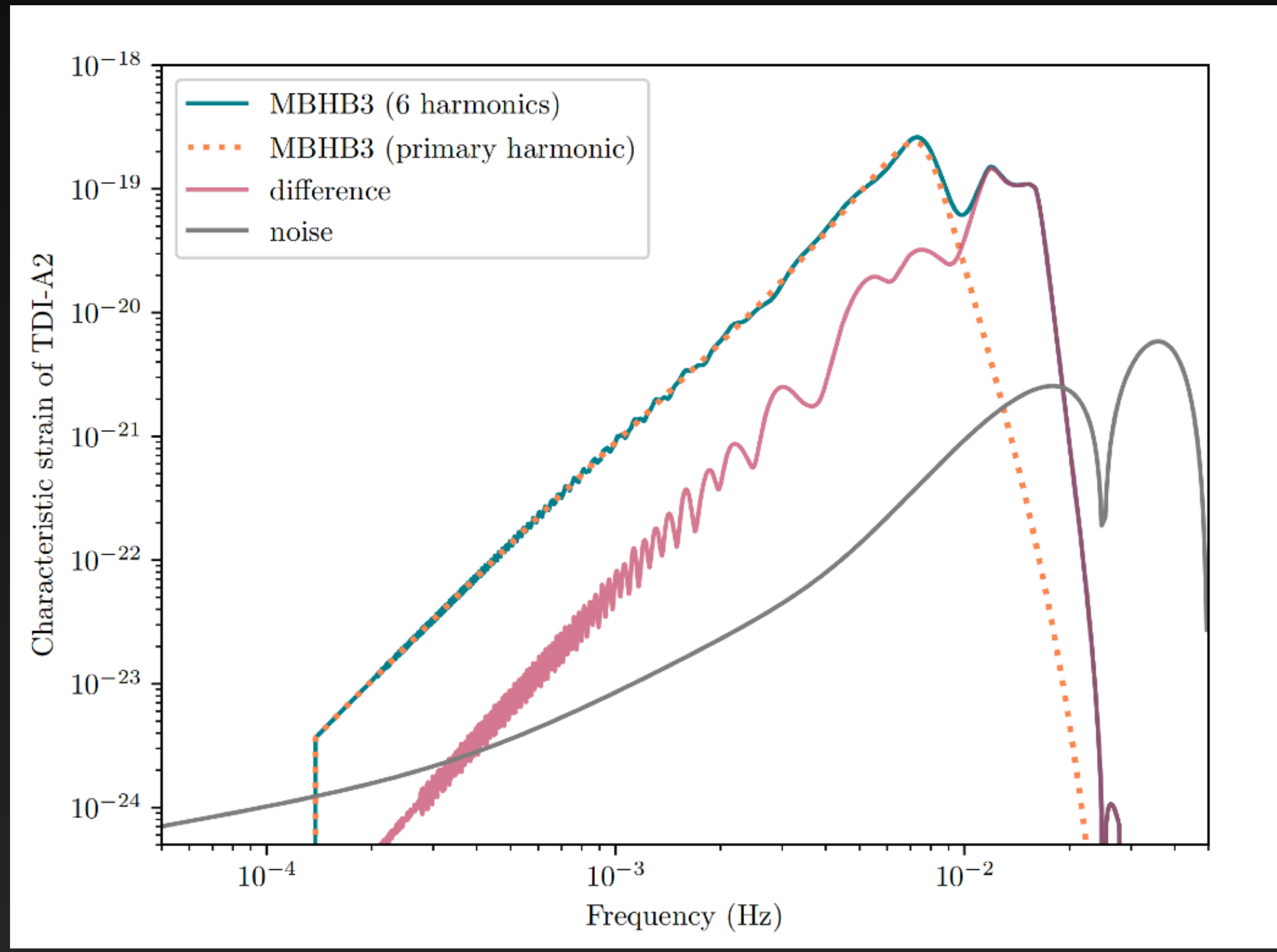
[5] C. Cutler et al, Phys. Rev. D 76, 104018, 2007

Template = Waveform + Projection + TDI

① Complexity of waveforms



Harmonics of MBHB waveforms[2]



MBHB3 of TDCII with & without eccentricity and high harmonics

Inclusion of high harmonics { Better sky localization; Break d_L-ι degeneracy } Multimessenger Cosmology

The "subdominant" features will be non-negligible

Publications on this topic:

Eccentricity:

[1] M. Garg et al, *Mon.Not.Roy.Astron.Soc.* 528 (2024) 3, 4176-4187

Higher harmonics:

[2] Y. Gong, Z. Cao, J. Zhao, L. Shao, Phys. Rev. D 108, 064046, 2023

[3] C. Pitte et al, Phys. Rev. D 108, 044053, 2023

[4] J.-D. Liu, W. -B. Han, Q. Yun, S.-C. Yang, *Mon.Not.Roy.Astron.Soc.* 532 (2024) 4, 4722-4728

Source environment and GW propagation:

[5] X. Chen, Handbook of Gravitational Wave Astronomy. Springer, Singapore.

[6] S.-J. Jin et al, Sci. China-Phys. Mech. Astron. 67, 220412 (2024)

EMRIs:

[7] P. Shen, Q. Cui, W.-B. Han, Phys. Rev. D 111, 024004, 2025

Template = Waveform + Projection + TDI

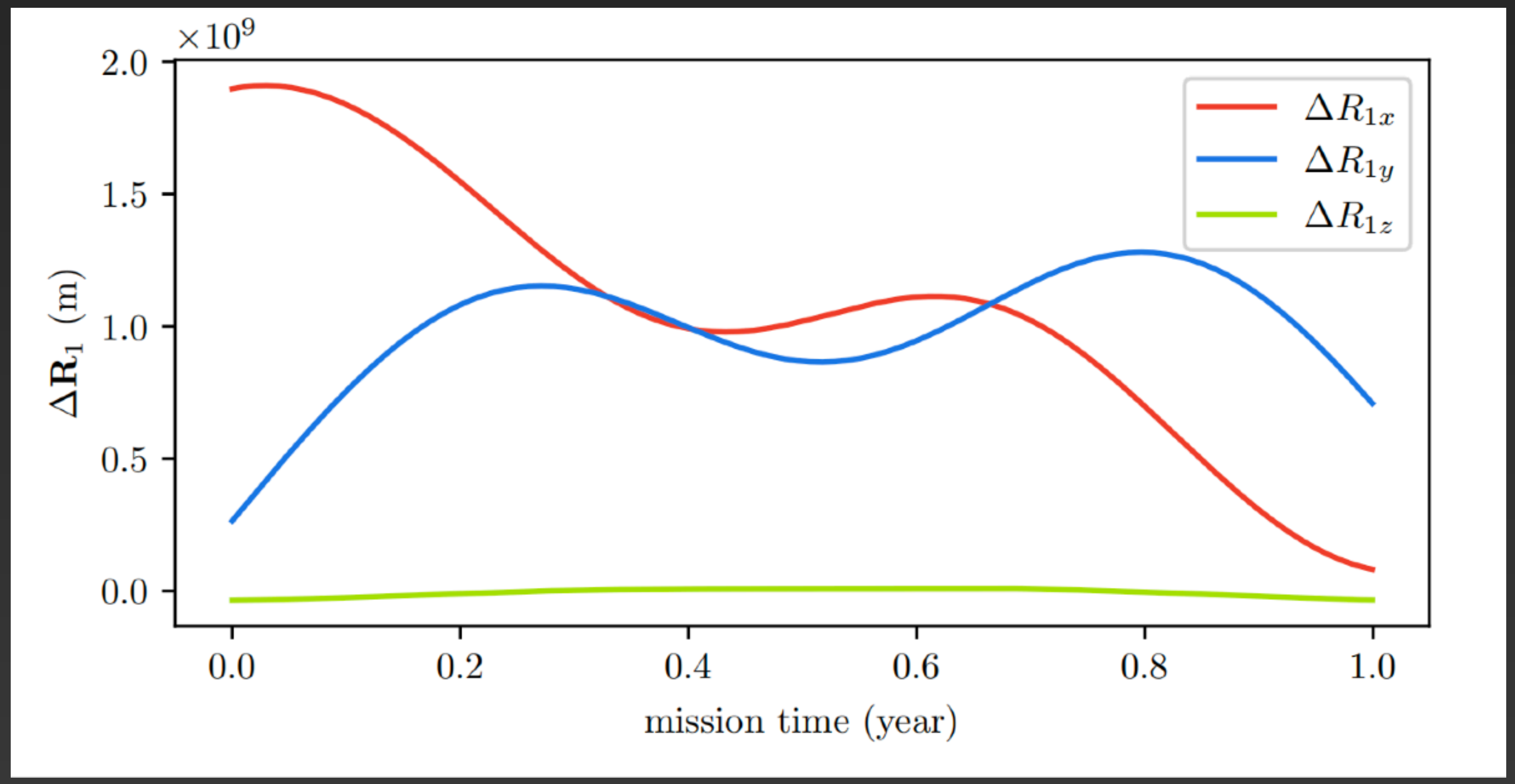
② Complexity in detector response

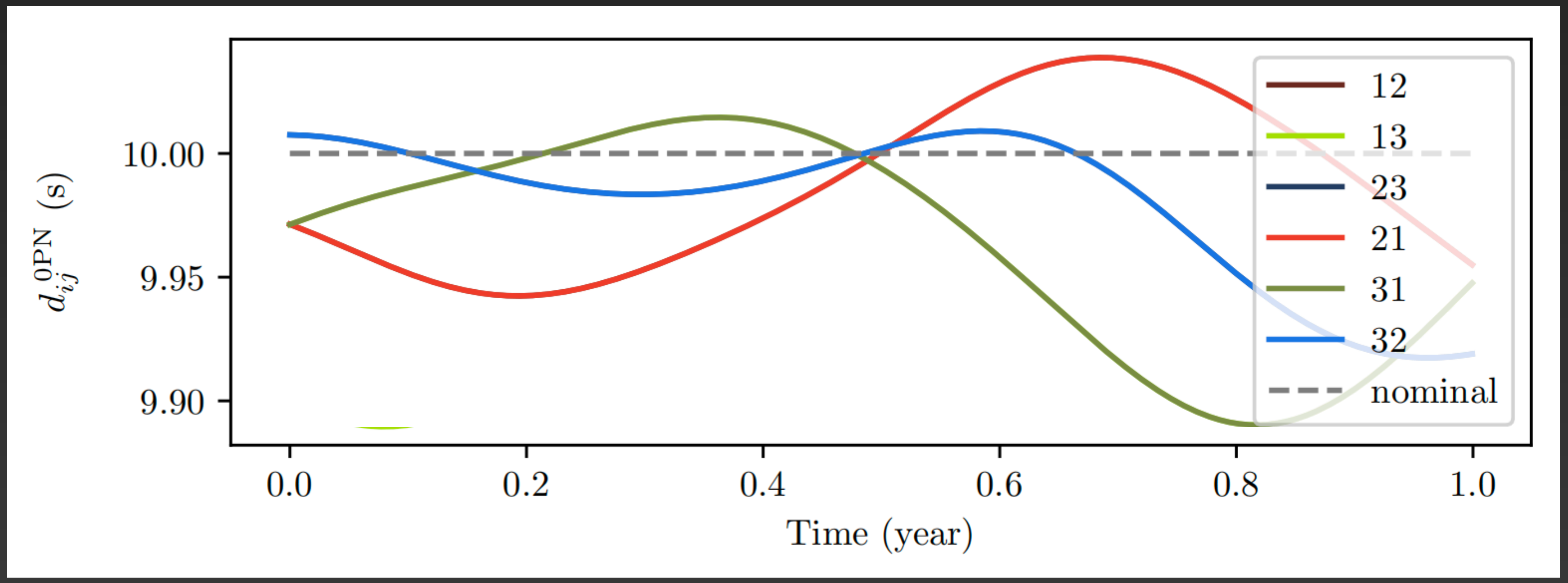
where orbit information enters

$$\eta_{ij}(t) \equiv \frac{\nu_{\rm receive} - \nu_{\rm send}}{\nu_{\rm send}} \approx \frac{1}{2\left[1 - \hat{\boldsymbol{k}} \cdot \boxed{\hat{\boldsymbol{n}}_{ij}(t)}\right]} \left[\Phi_{ij}\left(t - \frac{\boxed{d_{ij}(t)}}{c} - \frac{\hat{\boldsymbol{k}} \cdot \boxed{\boldsymbol{R}_j(t)}}{c}\right) - \Phi_{ij}\left(t - \frac{\hat{\boldsymbol{k}} \cdot \boxed{\boldsymbol{R}_i(t)}}{c}\right)\right], \quad \Phi_{ij}(t) = \boxed{\hat{\boldsymbol{n}}_{ij}(t) \otimes \hat{\boldsymbol{n}}_{ij}(t)} : \boldsymbol{h}(t)$$

$$X_2 = \left(1 - \boxed{\mathbf{D}_{121} - \mathbf{D}_{12131} + \mathbf{D}_{1312121}}\right)\left(\eta_{13} + \boxed{\mathbf{D}_{13}}\eta_{31}\right) - \left(1 - \boxed{\mathbf{D}_{131} - \mathbf{D}_{13121} + \mathbf{D}_{1213131}}\right)\left(\eta_{12} + \boxed{\mathbf{D}_{12}}\eta_{21}\right)$$

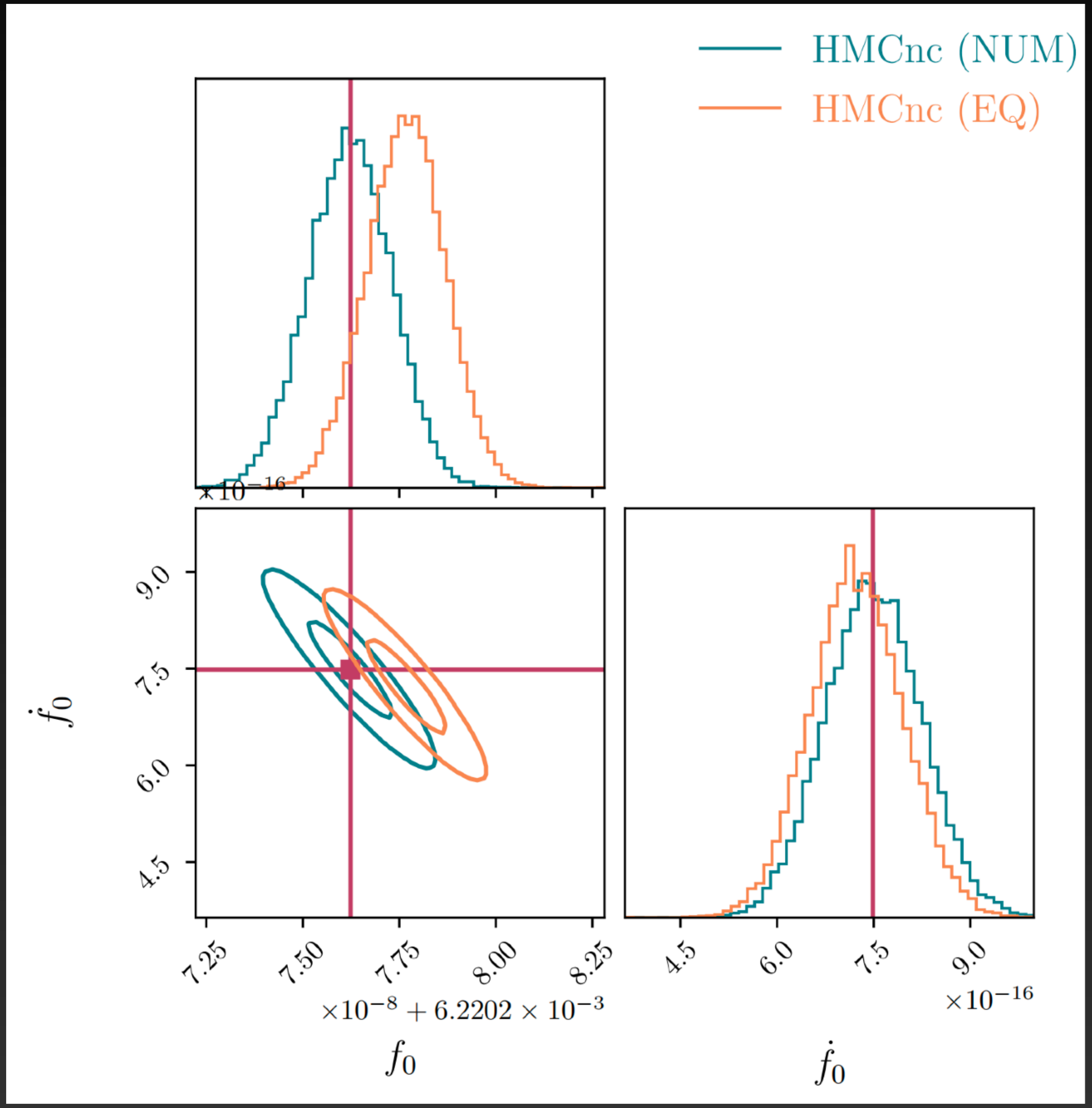
The difference between TDC orbit and its closest equal-arm orbit: $(\kappa_0, \alpha_0) = \mathrm{argmin}\left\{ \sum_{t,\mathrm{SC}} \left| \boldsymbol{R}^{\rm NUM}_{\rm SC}(t) - \boldsymbol{R}^{\rm EQ}_{\rm SC}(t; \kappa_0, \alpha_0)\right|^2 \right\}$



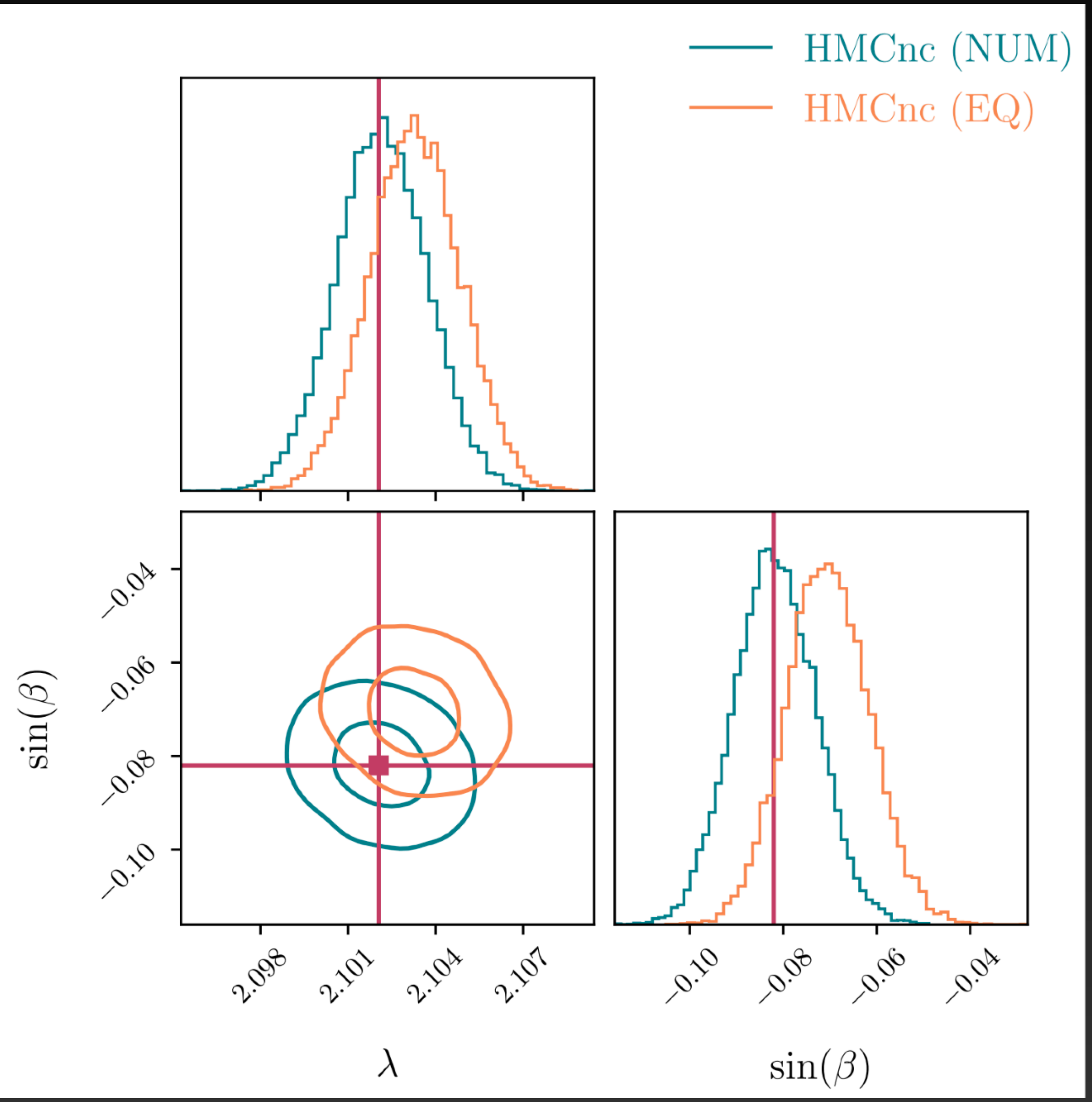


Test on verification binary HMCnc:

Simulation: **NUM orbit**; parameter estimation: **NUM orbit** vs. **EQ orbit**



Posterior of intrinsic parameters

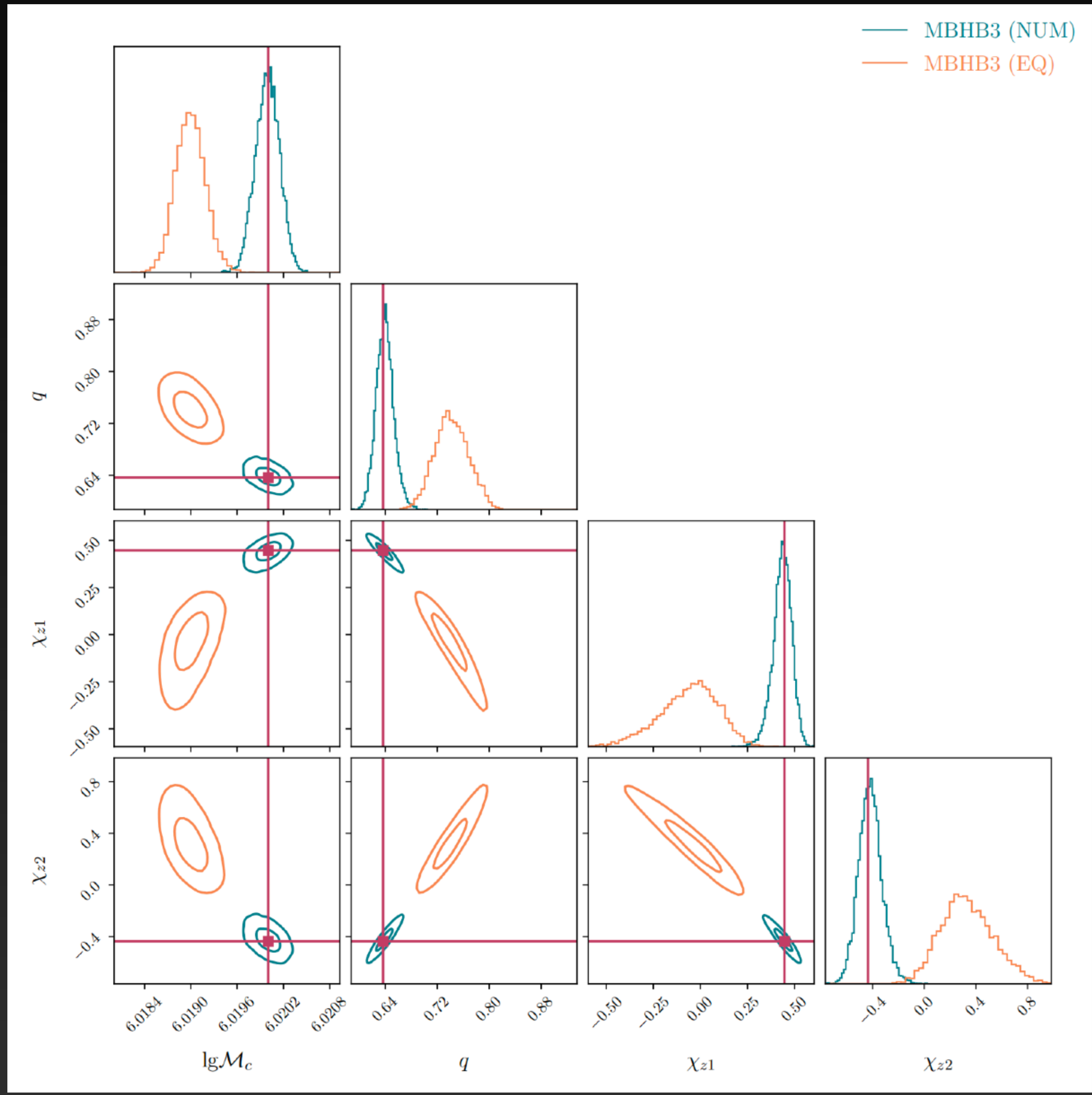


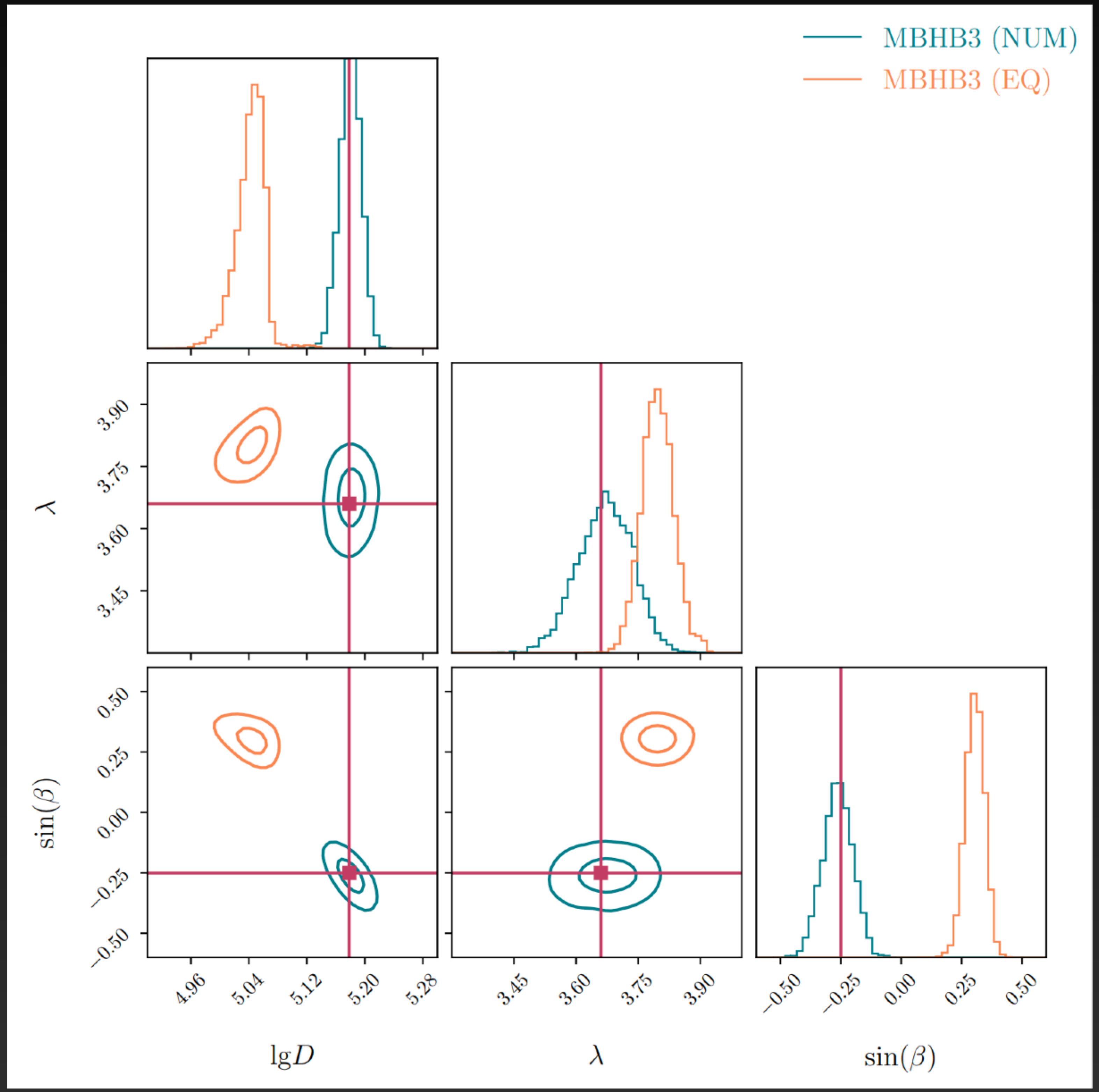
Posterior of sky locations

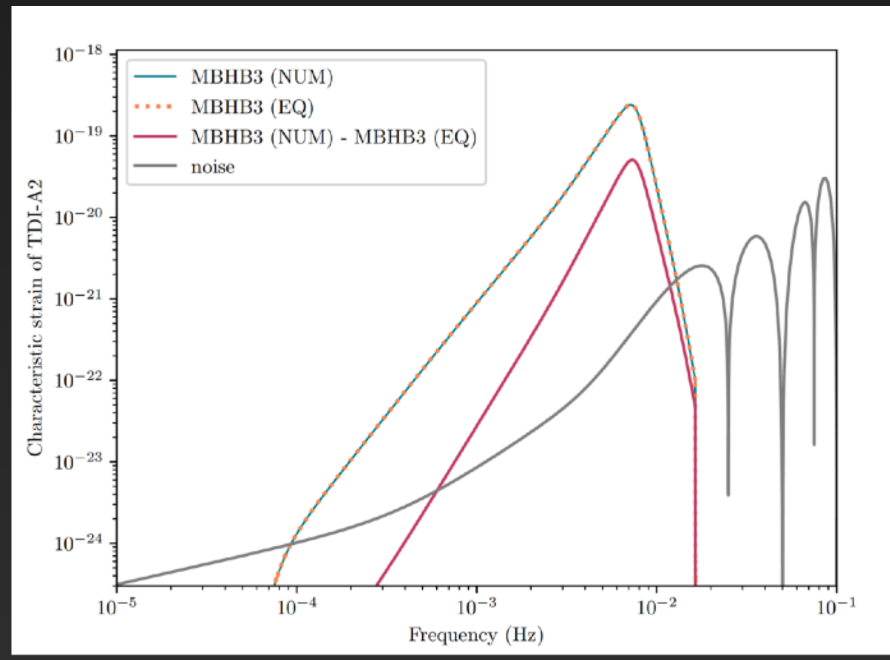
See a more catastrophic case in M. Katz, Phys. Rev. D 106, 103001 (2022)

Test on the 3rd MBHB of TDC:

Simulation: **NUM orbit**; parameter estimation: **NUM orbit** vs. **EQ orbit**







Posterior of intrinsic parameters

Posterior of 3D locations

Templates and data analysis tools must be applicable to numerical orbits

2. The complexity of noise modeling and processing

① Complexity due to the lack of noise knowledges

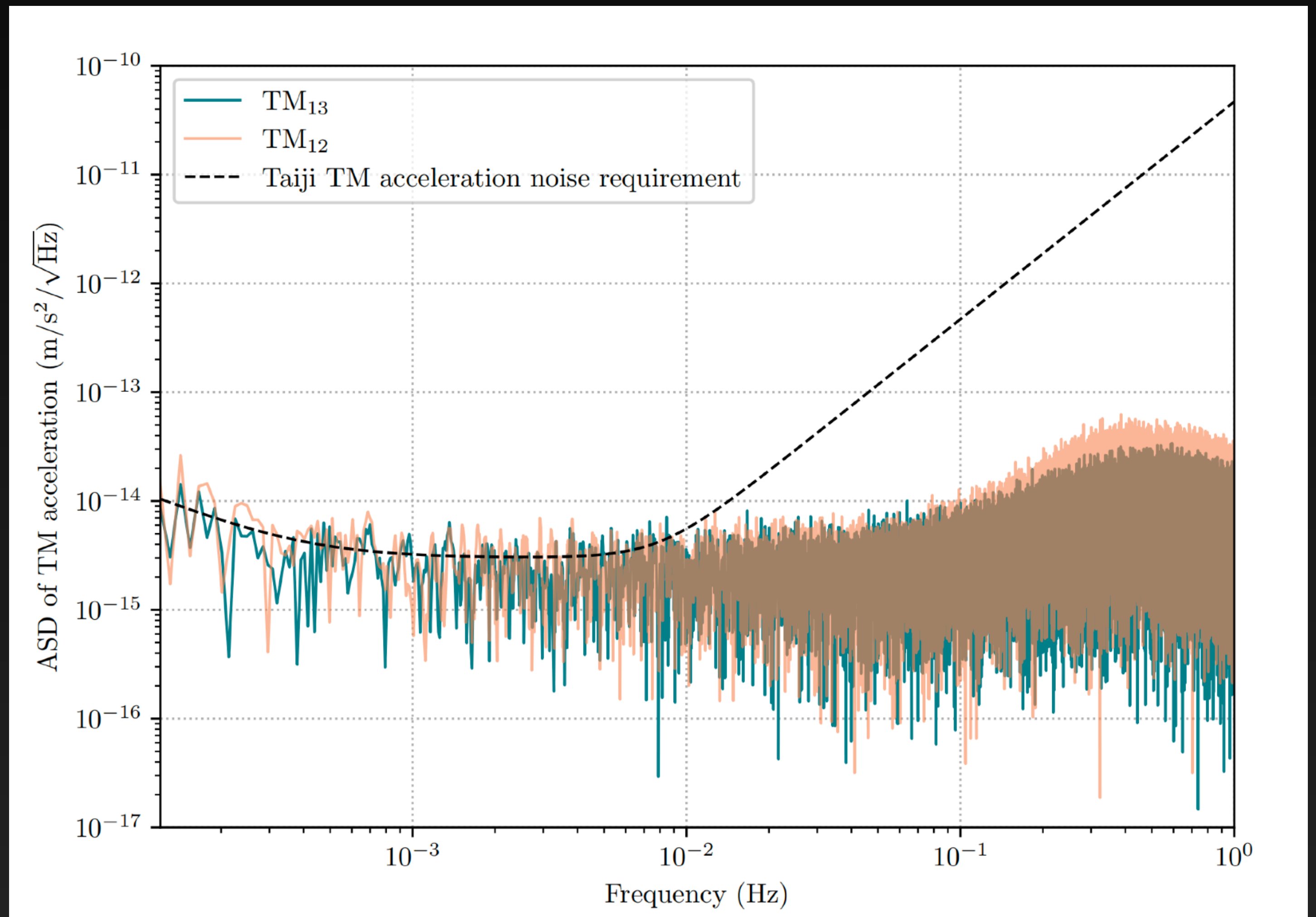
Matched filter

$$(d\,|\,h) \equiv 4\mathrm{Re}\int_{f_{\min}}^{f_{\max}} \frac{\tilde{d}(f)\tilde{h}^*(f)}{S(f)}df$$

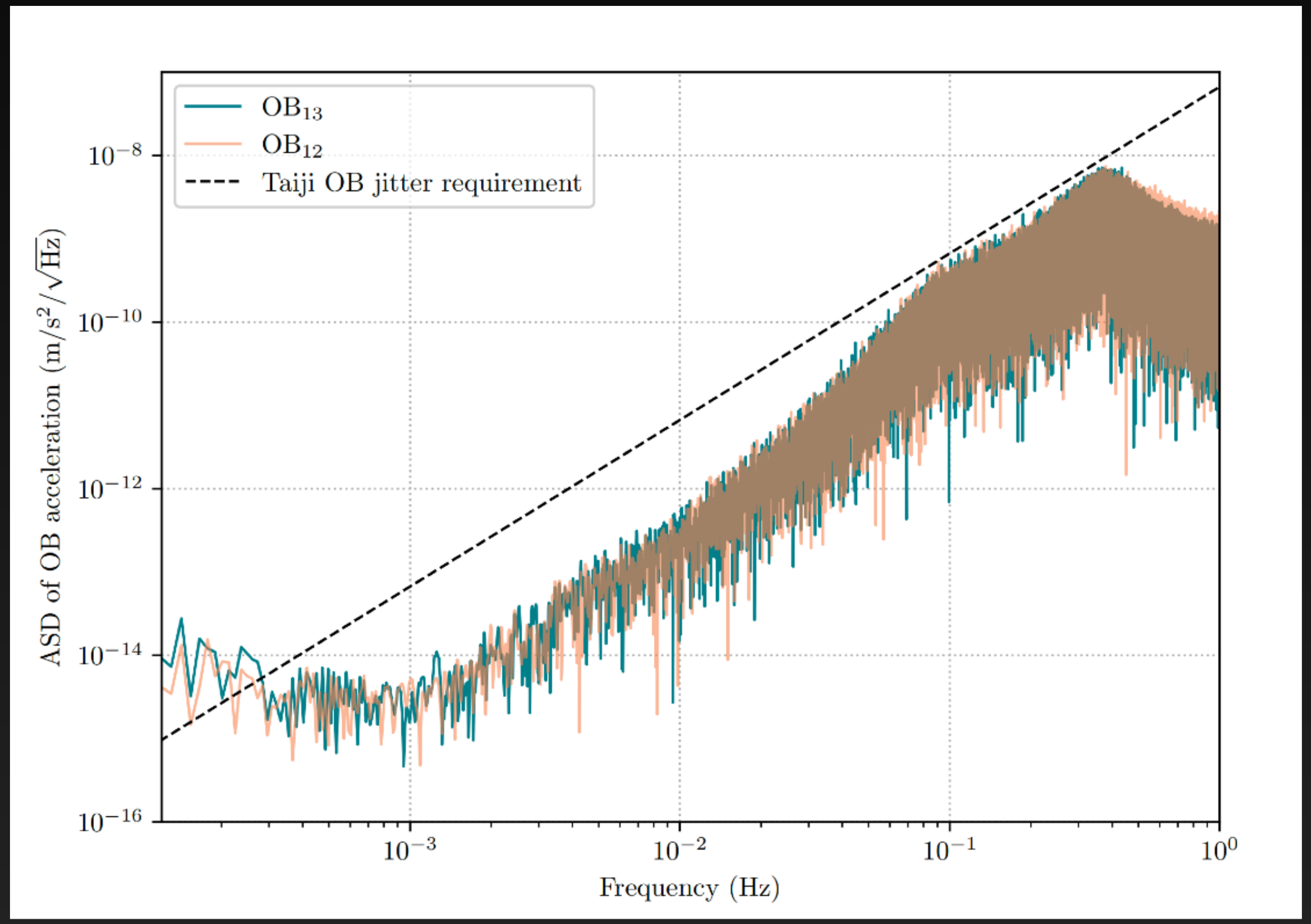
data

template

noise



Theoretical requirements VS simulated TM acceleration noises



Theoretical requirements VS simulated SC jitters

Noise models should be estimated along with signals

Publications on this topic:

The impact of noise knowledge: [1] M. Muratore et al, Phys. Rev. D 109, 042001, 2024

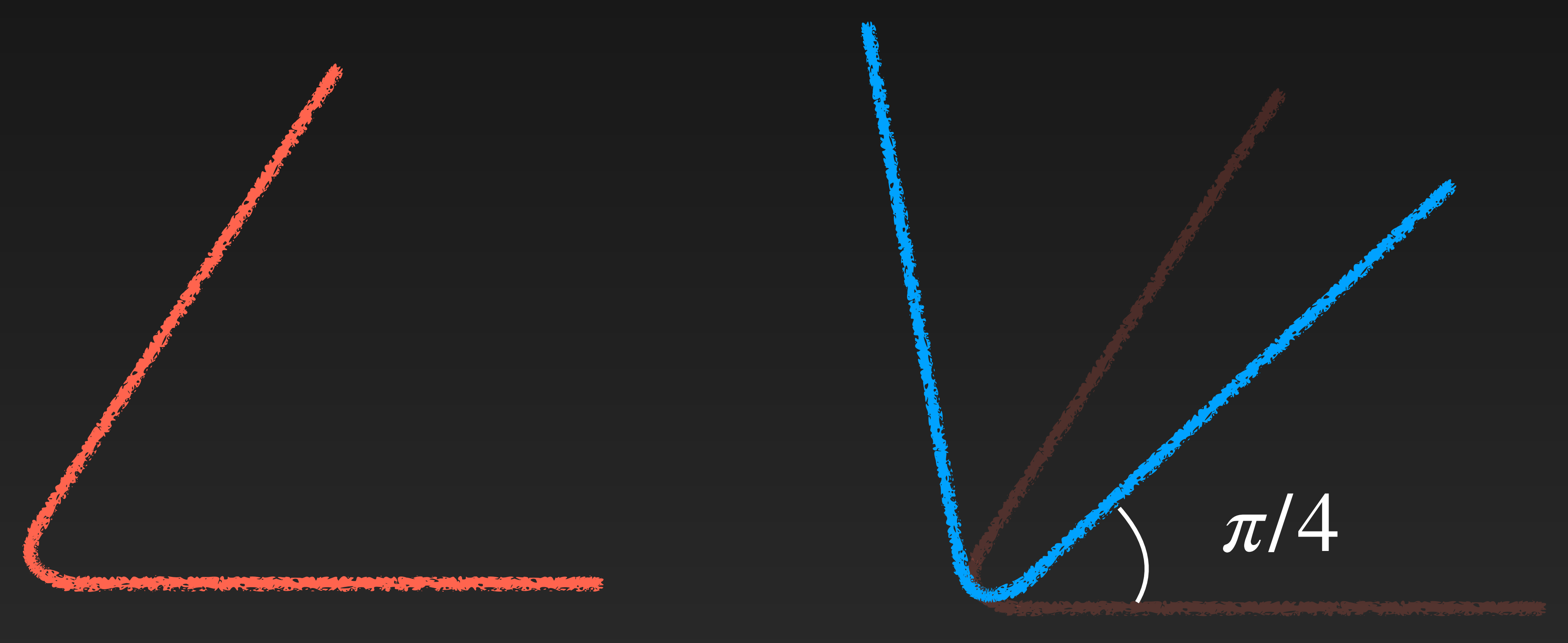
SGWB estimation under unknown noise shape: [2] Q. Baghi et al, JCAP04(2023)066

SGWB detection under unknown noise amplitude: [3] N Karnesis et al 2020 CQG 37 215017

Gaussian Process: [4] F. Pozzoli et al, Phys. Rev. D 109, 083029, 2024

② Complexity due to TDI in realistic scenarios (unequal and time-varying armlengths)

Michelson A, E, T channels under the low-frequency, equal-arm, equal-noise approximations:



Michelson-A

Michelson-E

- **Noise orthogonal:**

$$\langle \cdot | \cdot \rangle_{\rm total} = \sum_{i \in AET} \langle \cdot | \cdot \rangle_i$$

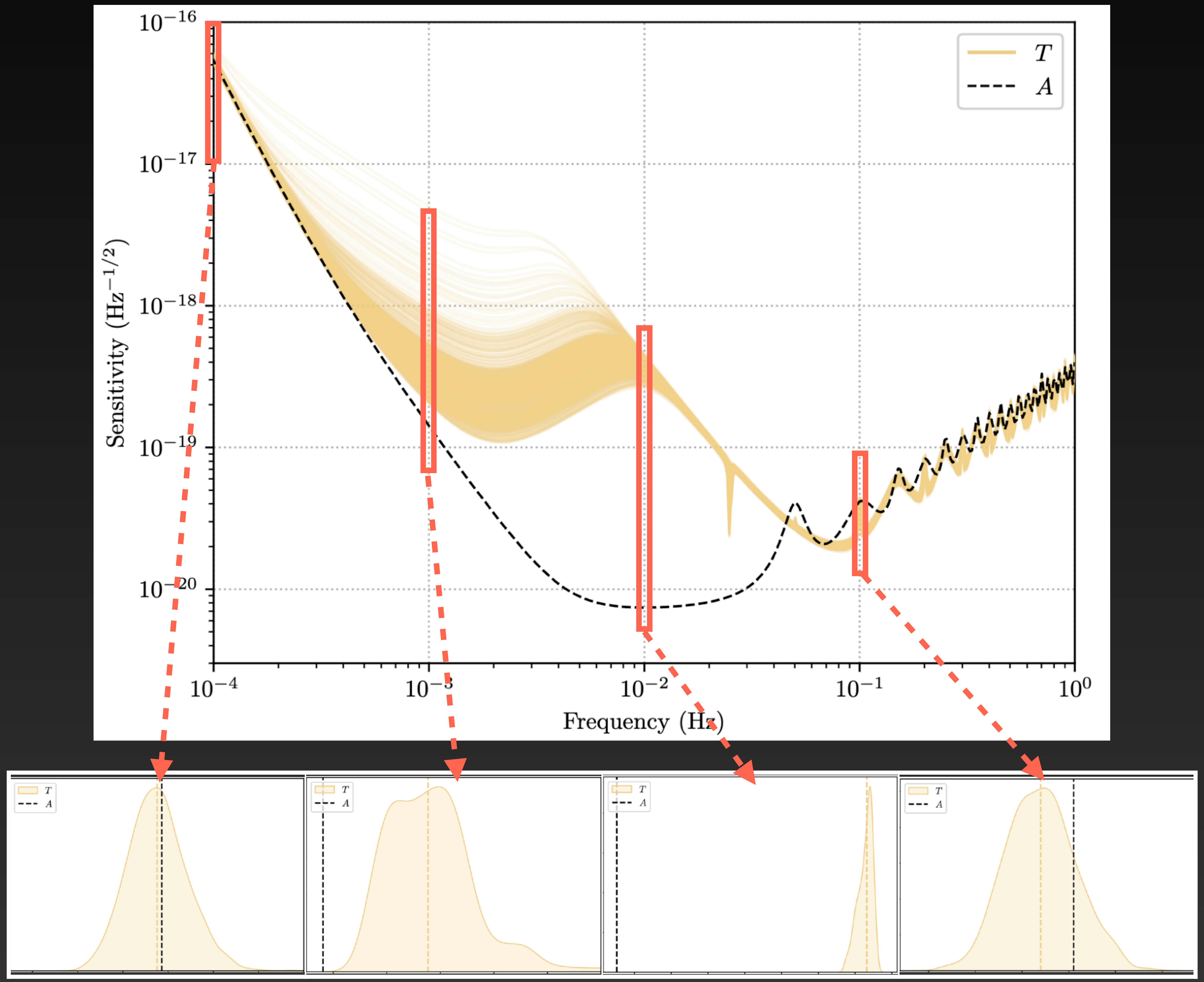
- **2 signal channels:**

  A, E, ~ 2 interferometers rotated by $\pi/4$

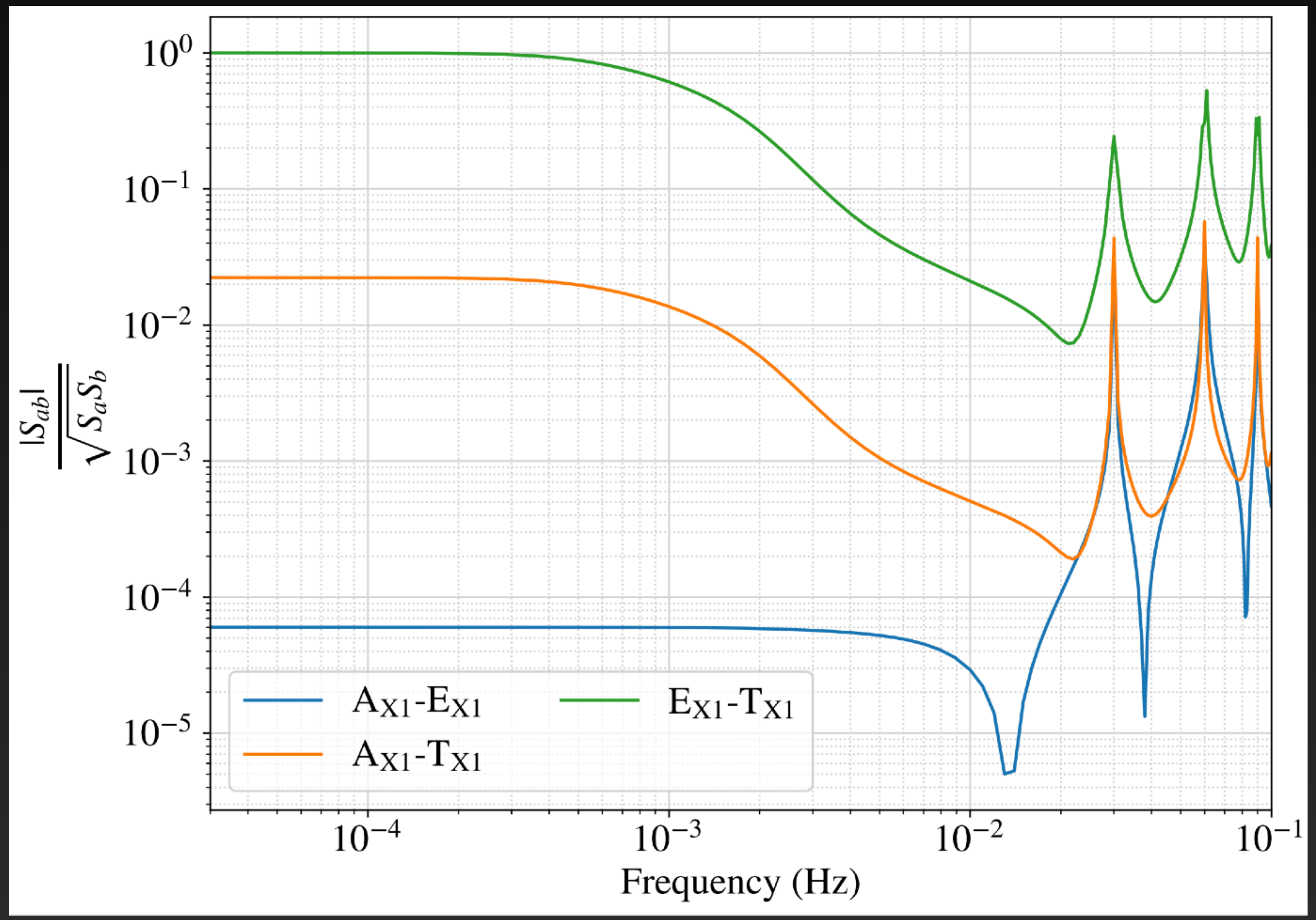
- **1 noise channel:**

  T, zero GW response, important for noise estimation and SGWB detection

Michelson A, E, T channels in realistic detection:



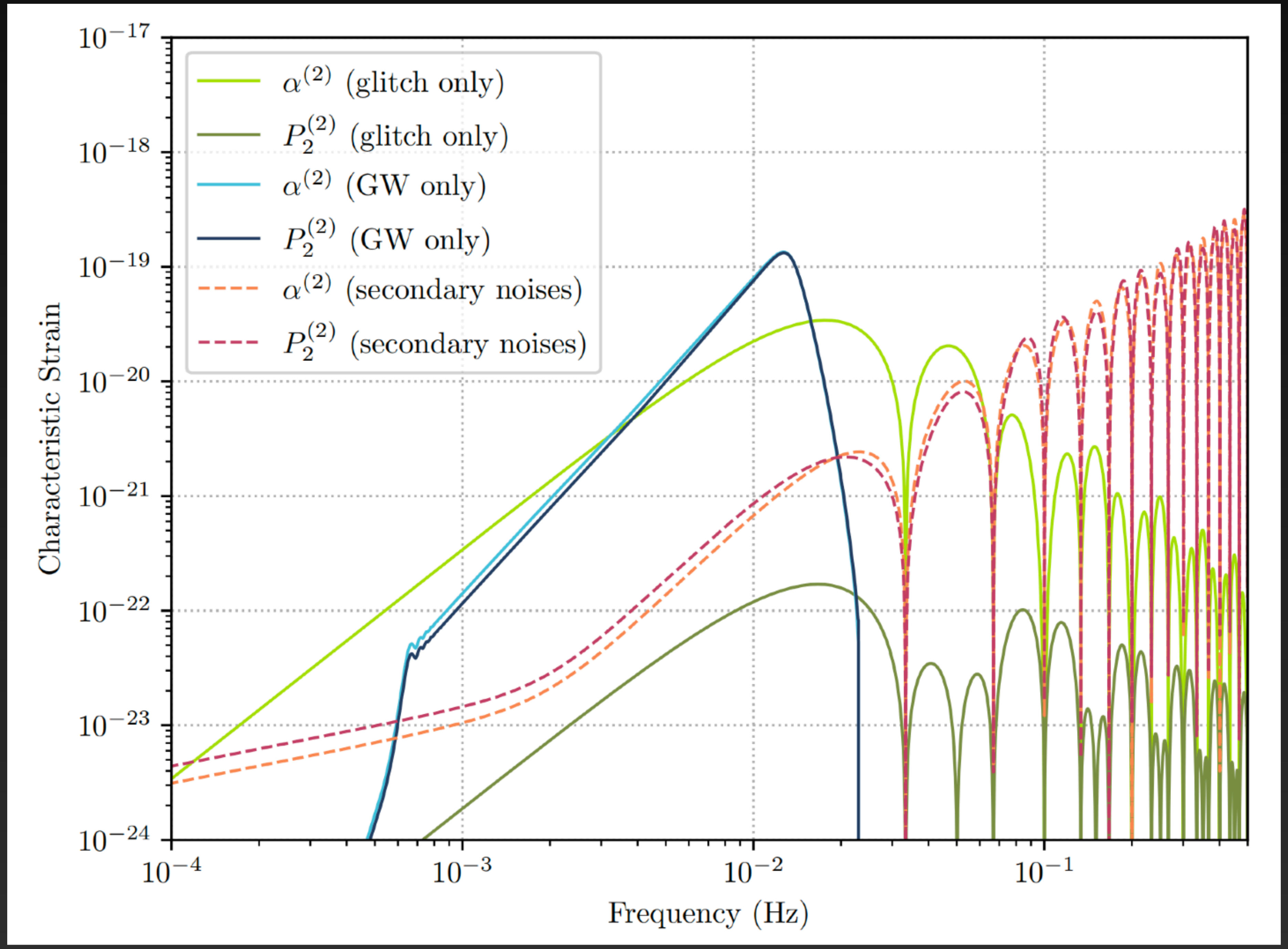
The sensitivity of T channel during 1 year



The correlation among AET channels [6]

Michelson-AET might not be the optimal choices, there are alternatives:

For example:



The TM glitch suppressing channel [11]

Publications on this topic:

The studies of LISA collaboration:

[1] O. Hartwig et al, Phys. Rev. D 105, 062006, 2022

[2] M. Muratore et al, Phys. Rev. D 107, 082004, 2023

[3] O. Hartwig et al, Phys. Rev. D 107, 123531, 2023

The studies of G. Wang:

[4] G. Wang et al, 2023 Phys. Scr. 98 075005

[5] G. Wang et al, Phys. Rev. D 106, 044054, 2022

[6] G. Wang, Phys. Rev. D 110, 042005, 2024

[7] G. Wang, Phys. Rev. D 110, 064085, 2024

[8] G. Wang, arXiv:2502.03983

Other TDI schemes with specific functions:

[9] P.-P. Wang et al, Results Phys., Vol. 58, 2024, 107481

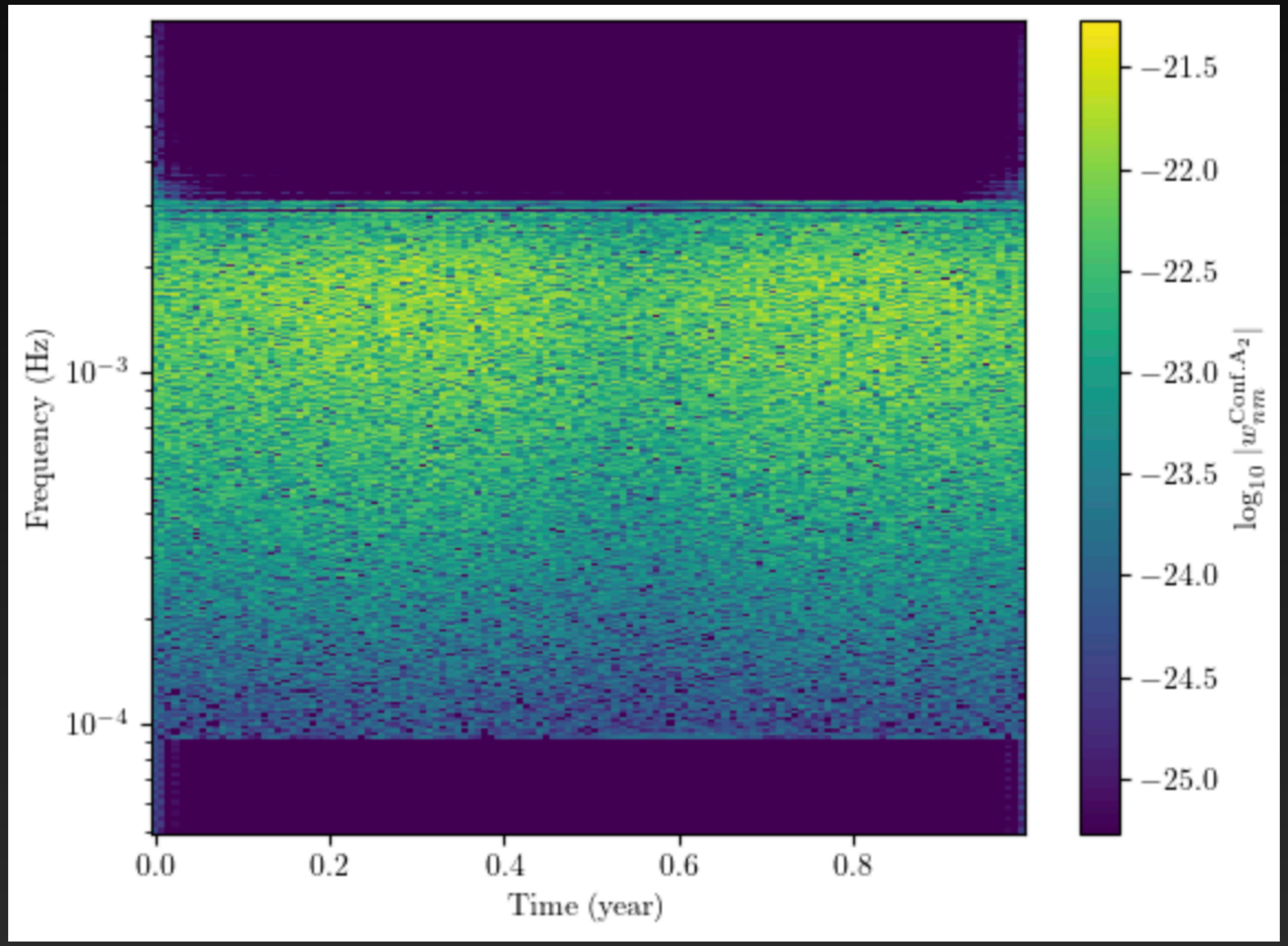
[10] Y.-J. Tan et al, Phys. Rev. D 111, 024011, 2025

[11] P. Wu et al, Opt.Express 32 (2024) 24, 43249-43263

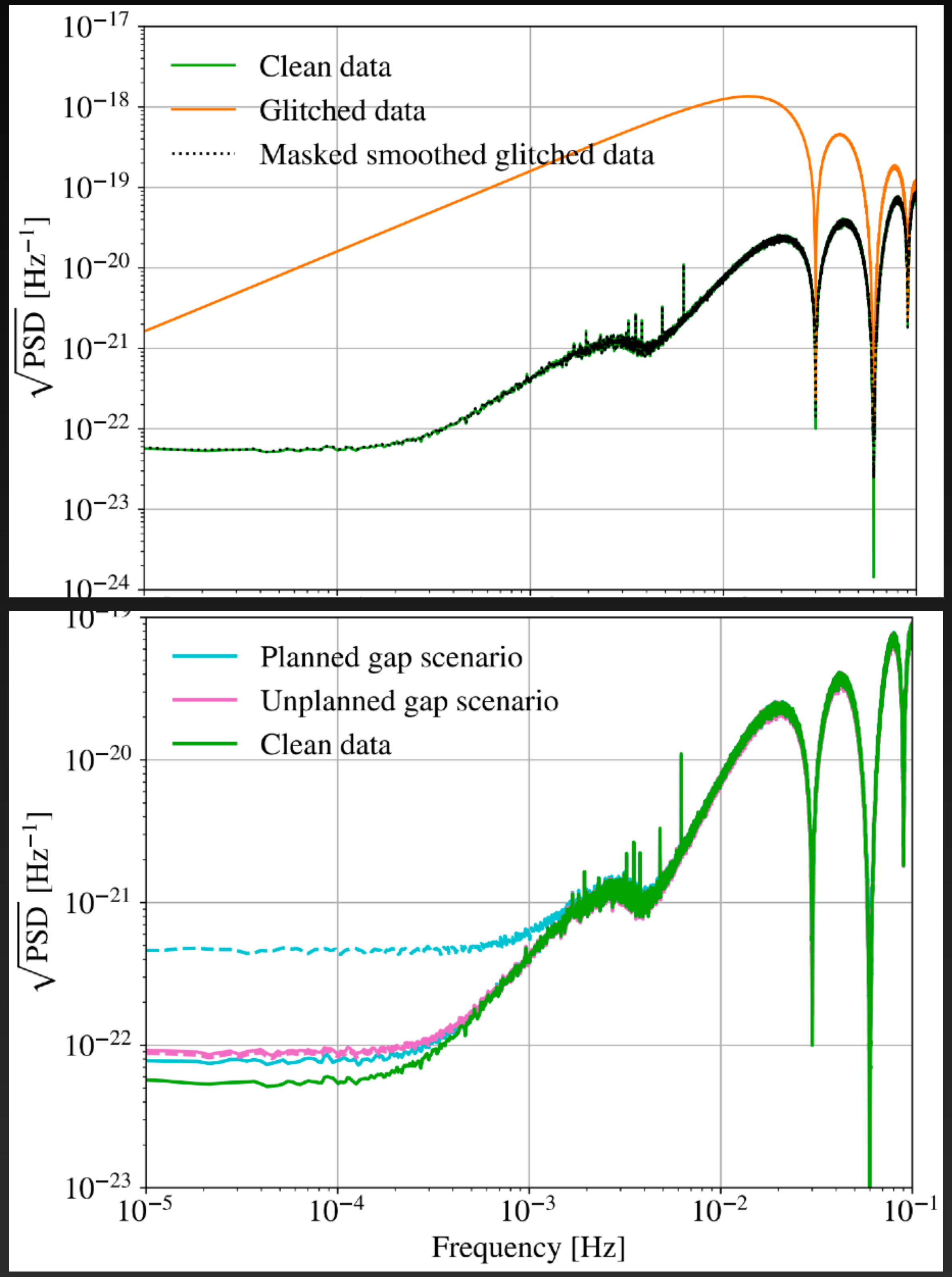
… …

Which are the best among them?

③ Complexity due to non-stationary noises and anomalies



The time-varying confusion foreground (shown in wavelet domain)



The increases of noise levels caused by glitches and gaps [1]

Publications on this topic:

Glitch and gap:

[1] E. Castelli et al, arXiv:2411.13402

[2] Y. Xu et al, Phys. Lett. B, Vol. 858, 2024, 139016

Gap:

[3] O. Burke et al, arXiv:2502.17426

[4] A. Seoane et al, Gen Relativ Gravit 54, 3 (2022)

[5] K. Dey et al, Phys. Rev. D 104, 044035 (2021)

[6] Q. Baghi et al, Phys. Rev. D 100, 022003, 2019

[7] R. Mao et al, Phys. Rev. D 111, 024067, 2025

Glitch:

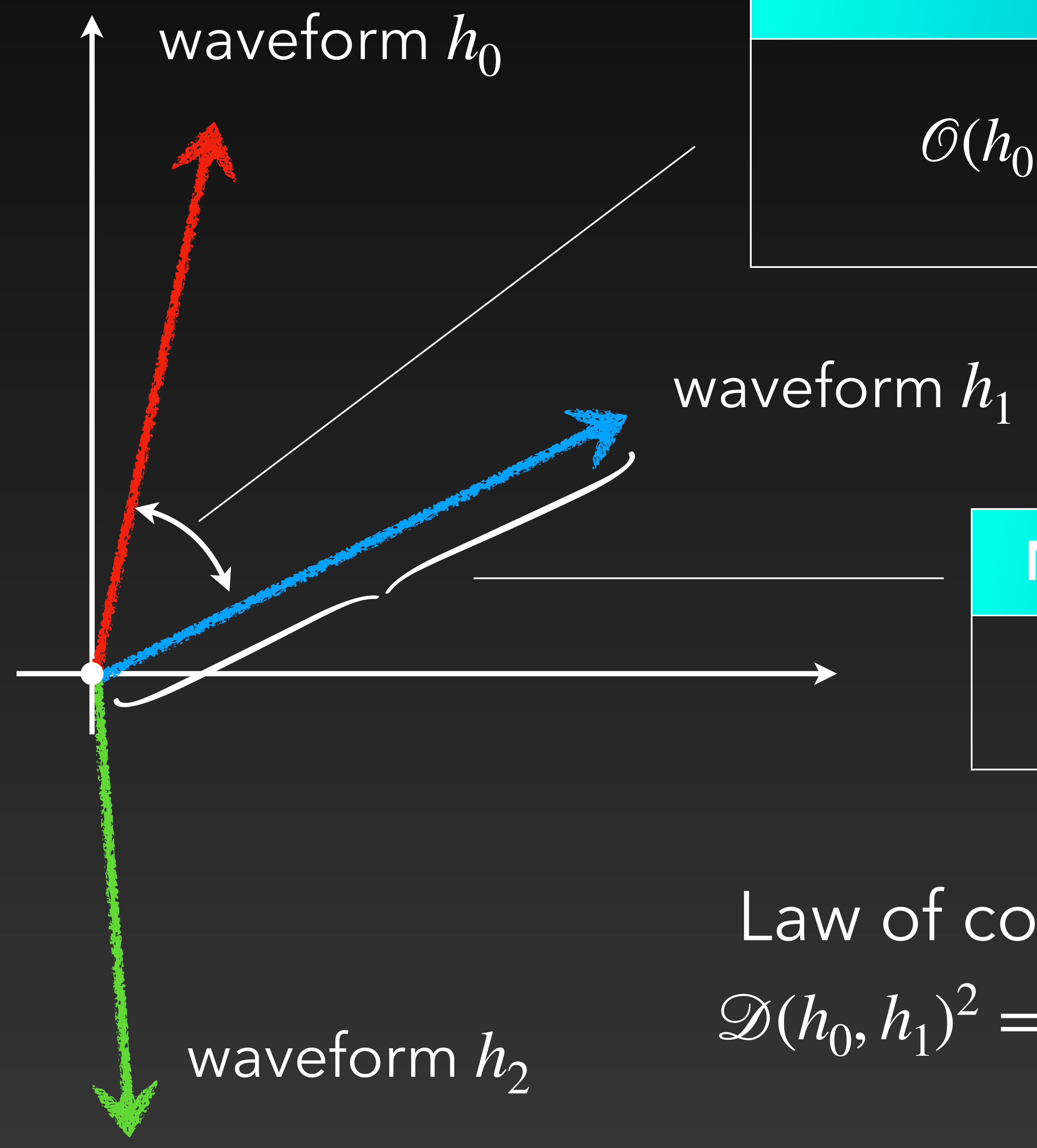
[8] A. Spadaro et al, Phys. Rev. D 108, 123029, 2023

[9] Q. Baghi et al, Phys. Rev. D 105, 042002, 2022

Theories and methods based on Gaussian stationary noises need reconsideration

3. Overlap of numerical signals

① Decrease in detectability



Angle: overlap

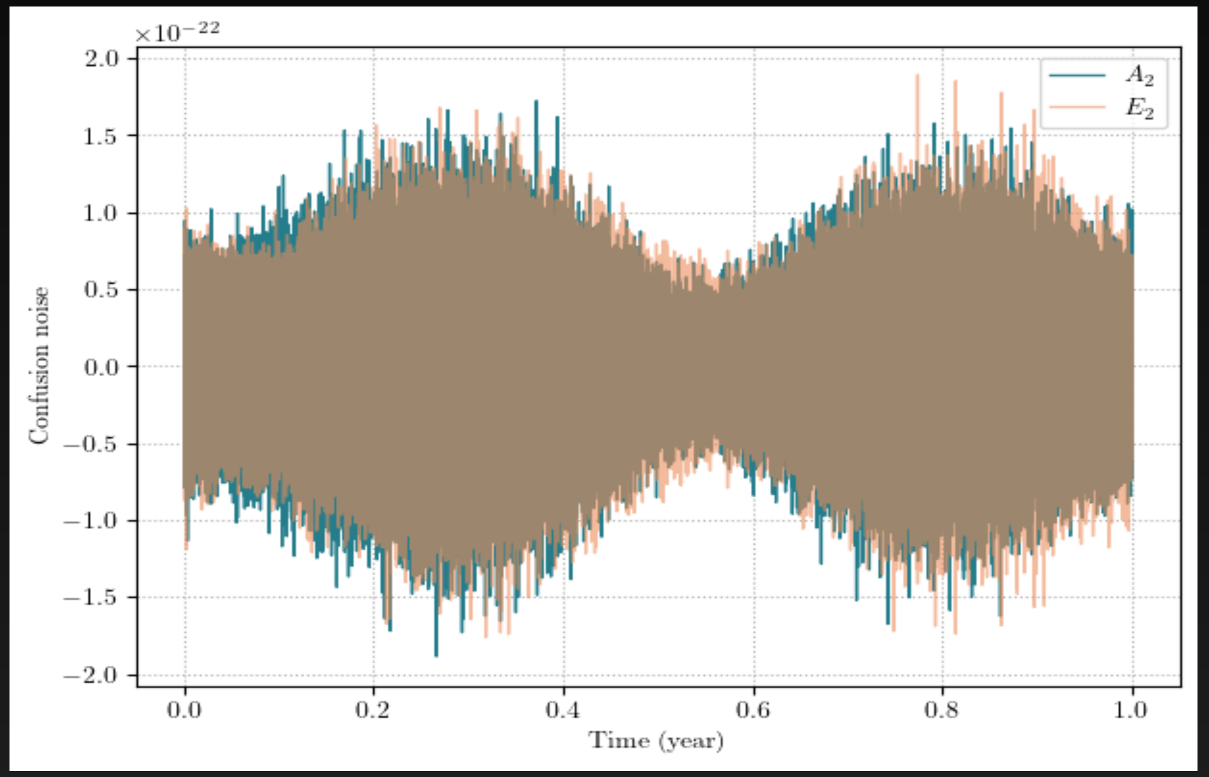
$$\mathcal{O}(h_0, h_1) = \frac{(h_0 | h_1)}{\sqrt{(h_0 | h_0)(h_1 | h_1)}}$$

Norm: signal-to-noise ratio

$$\rho = \sqrt{(h_1 | h_1)}$$

Law of cosines:

$$\mathcal{D}(h_0, h_1)^2 = \rho_0^2 + \rho_1^2 - 2\rho_0\rho_1\mathcal{O}(h_0, h_1)$$



Overlapped 10^7 GB signals

$$\rho(h_0),\ \rho(h_2) > \rho_{\mathrm{threshold}}$$

$$\rho(h_0 + h_2) < \rho_{\mathrm{threshold}}$$

Both h0 and h2 are undetectable

② The "global fit" challenge

Noise dominant ➡ Signal dominant

The full likelihood (assuming known noise model and known source numbers):

$$\ln\mathscr{L} = -\frac{1}{2}\left(d - h_{\text{total}} \,|\, d - h_{\text{total}}\right)$$

$$= -\frac{1}{2}\left(\sum_i h_i \middle| \sum_j h_j\right) + \left(\sum_i h_i \middle| d\right) + \text{const}\,.$$

$$= -\frac{1}{2}\sum_i \left(h_i \,|\, h_i\right) \boxed{-\sum_{i<j}\left(h_j \,|\, h_i\right)} + \sum_i \left(h_i^{\text{true}} \middle| h_i\right) \boxed{+2\sum_{i<j}\left(h_j^{\text{true}} \middle| h_i\right)} + \sum_i \left(n \,|\, h_i\right) + \text{const}\,.$$

correlation correlation

$$\neq \sum_i \ln\mathscr{L}_i$$

Maxima for each source ≠ global maxima ➡ Fitting all parameters simultaneously
High dimensional & trans-dimensional search

Publications on this topic:

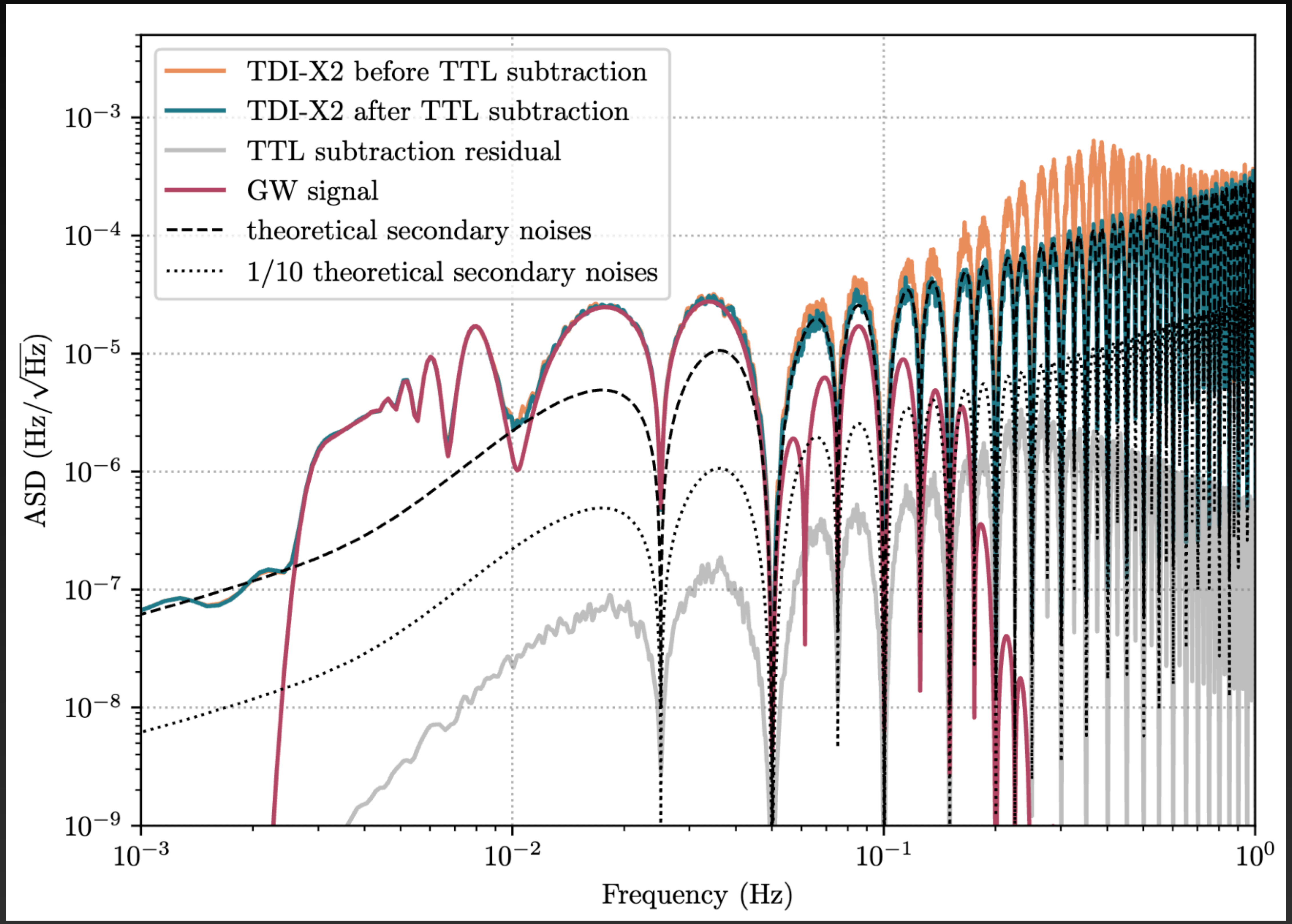
[1] T. B. Littenberg et al. Phys. Rev. D 107, 063004
[2] S. H. Strub et al. Phys. Rev. D 110, 024005
[3] M. Katz et al. Phys. Rev. D 111, 024060

4. The correlation between pre-processing and scientific analysis

① Noise characterization

② TTL noise suppression

③ Inter-SC ranging

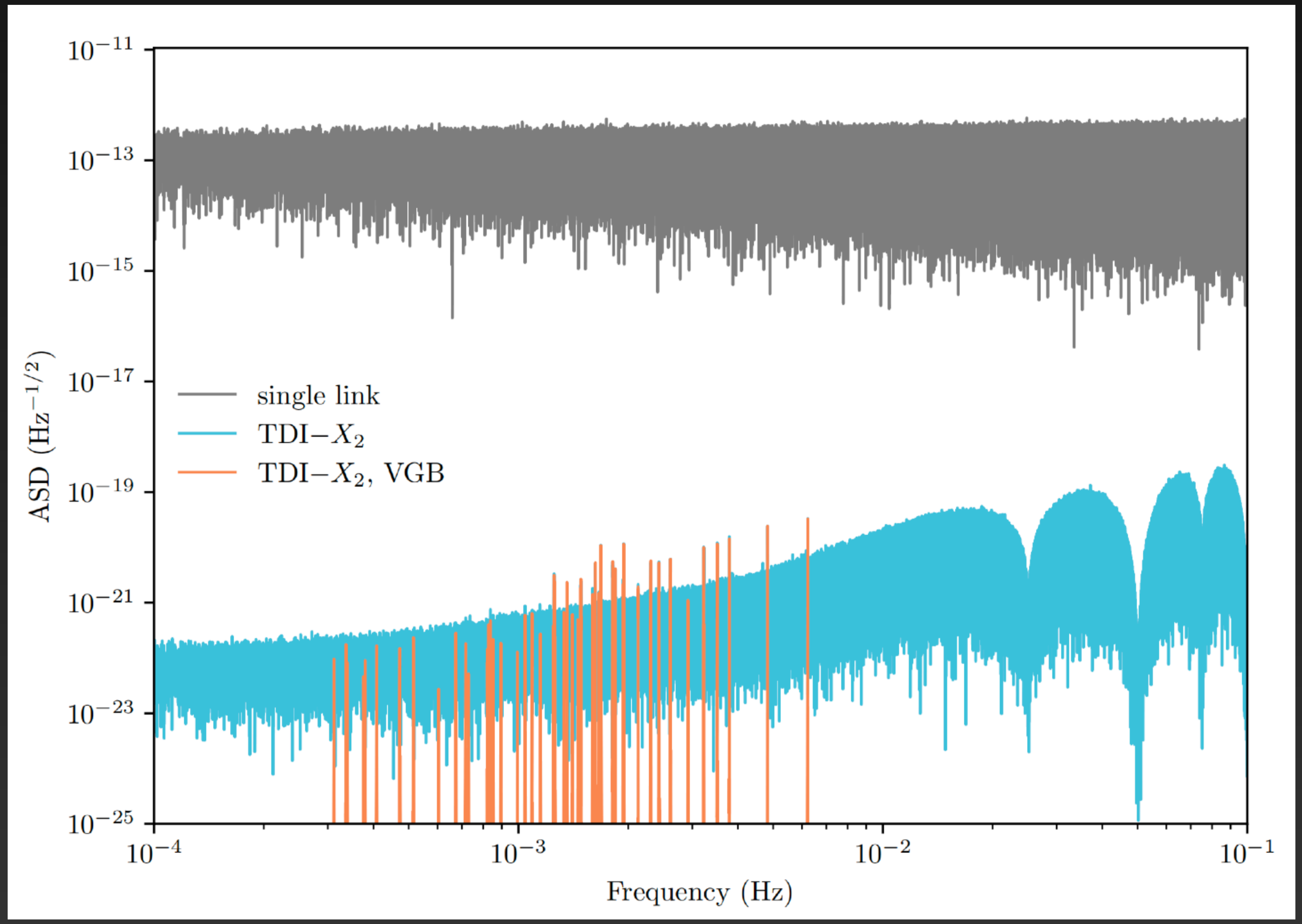


TTL noise suppression in the presence of GWs

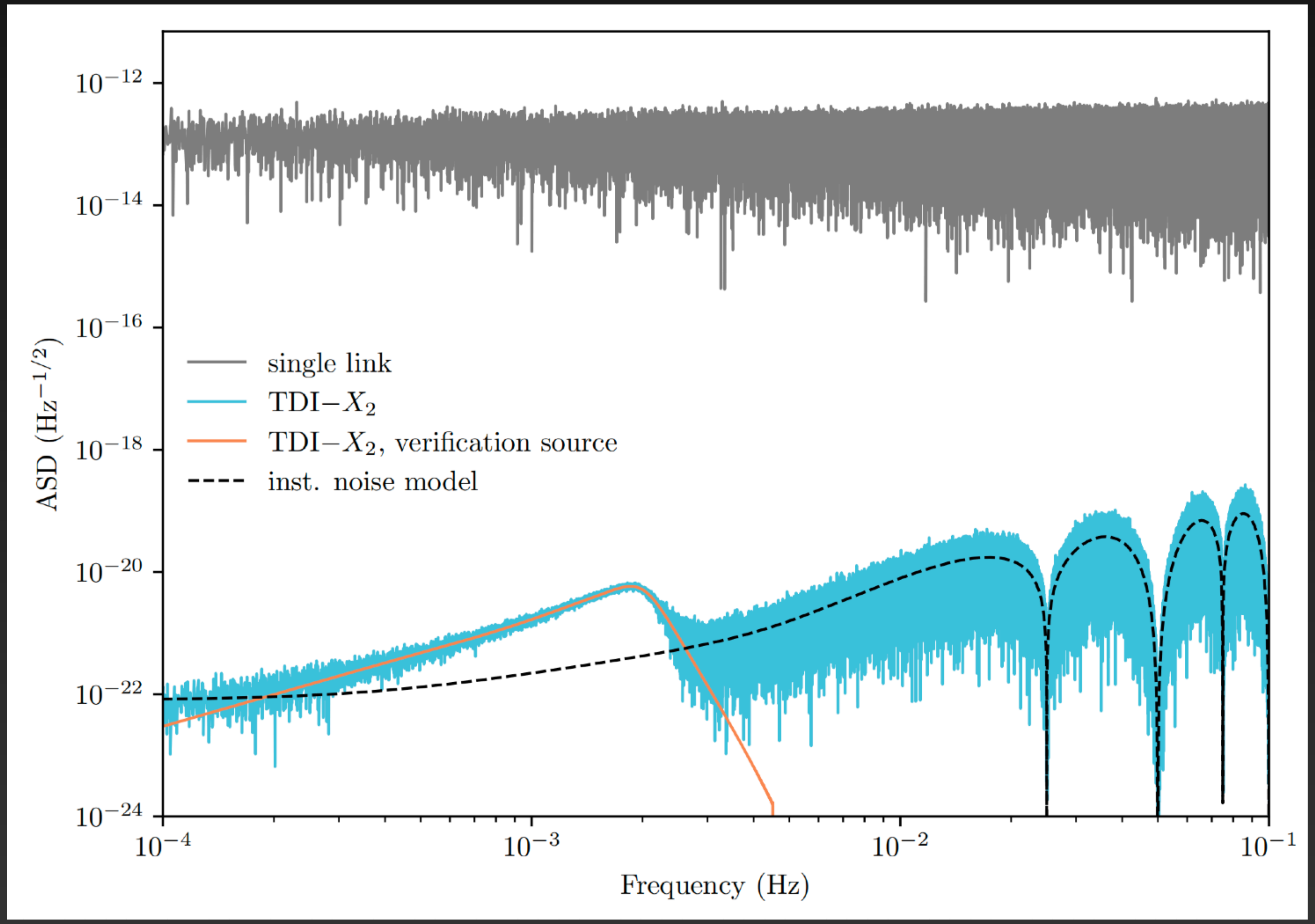
1. Basic Concepts
2. Mock Data Pipeline
3. Challenges
4. **The TDC Datasets**
5. Todos
6. Triangle
7. Tests and Verifications
8. Take-away messages

Challenge 0: The "Verification" Datasets

Target	Verify the modeling of signal responses and noises & prepare for the analysis of subsequent datasets. Simple demonstrations are provided.
Dataset 0.1/0.2	55 VGBs / 1 MBHB (PhenomD) + inst. noises



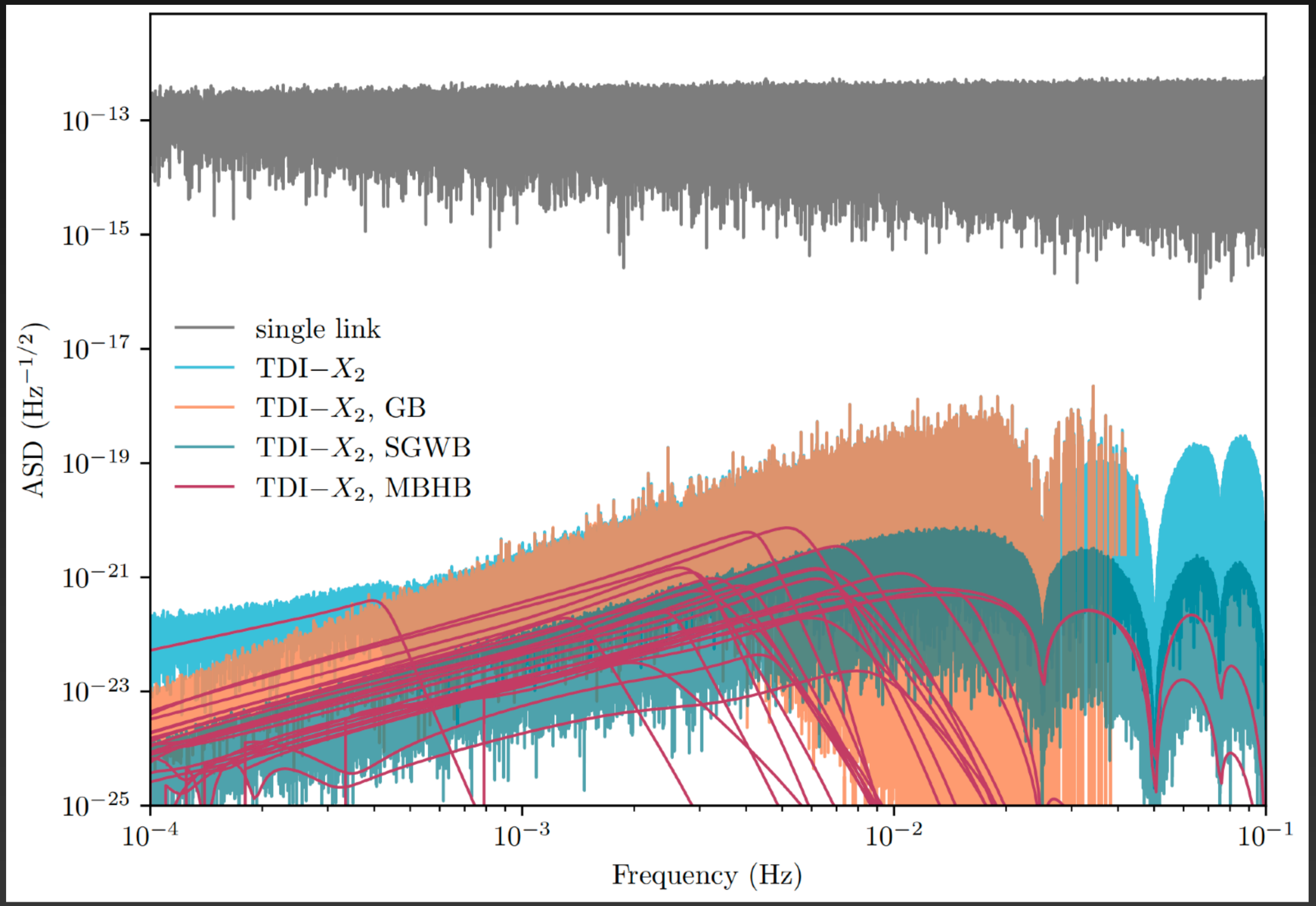
Dataset 0.1 in frequency domain (ASD)



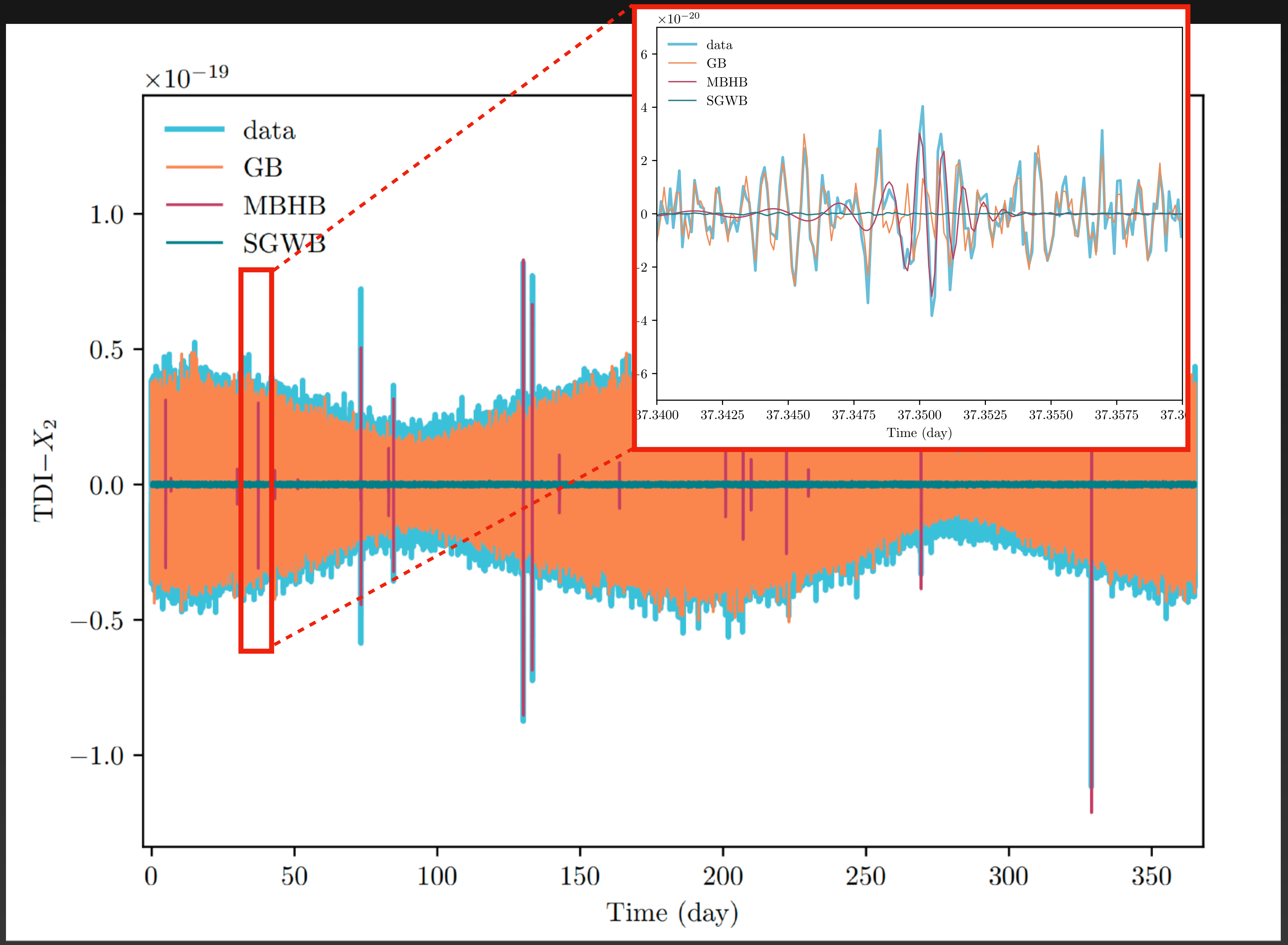
Dataset 0.2 in frequency domain (ASD)

Challenge 1: The "Global Fit" Dataset

Target	Address the global fit challenge under numerical orbit, more overlapping populations, unknown noise models, SGWB included for the 1st time, based on the "verified" waveforms
Dataset 1.1	25 MBHBs (PhenomD) + 4.5e7 GBs + SGWB + inst. noise



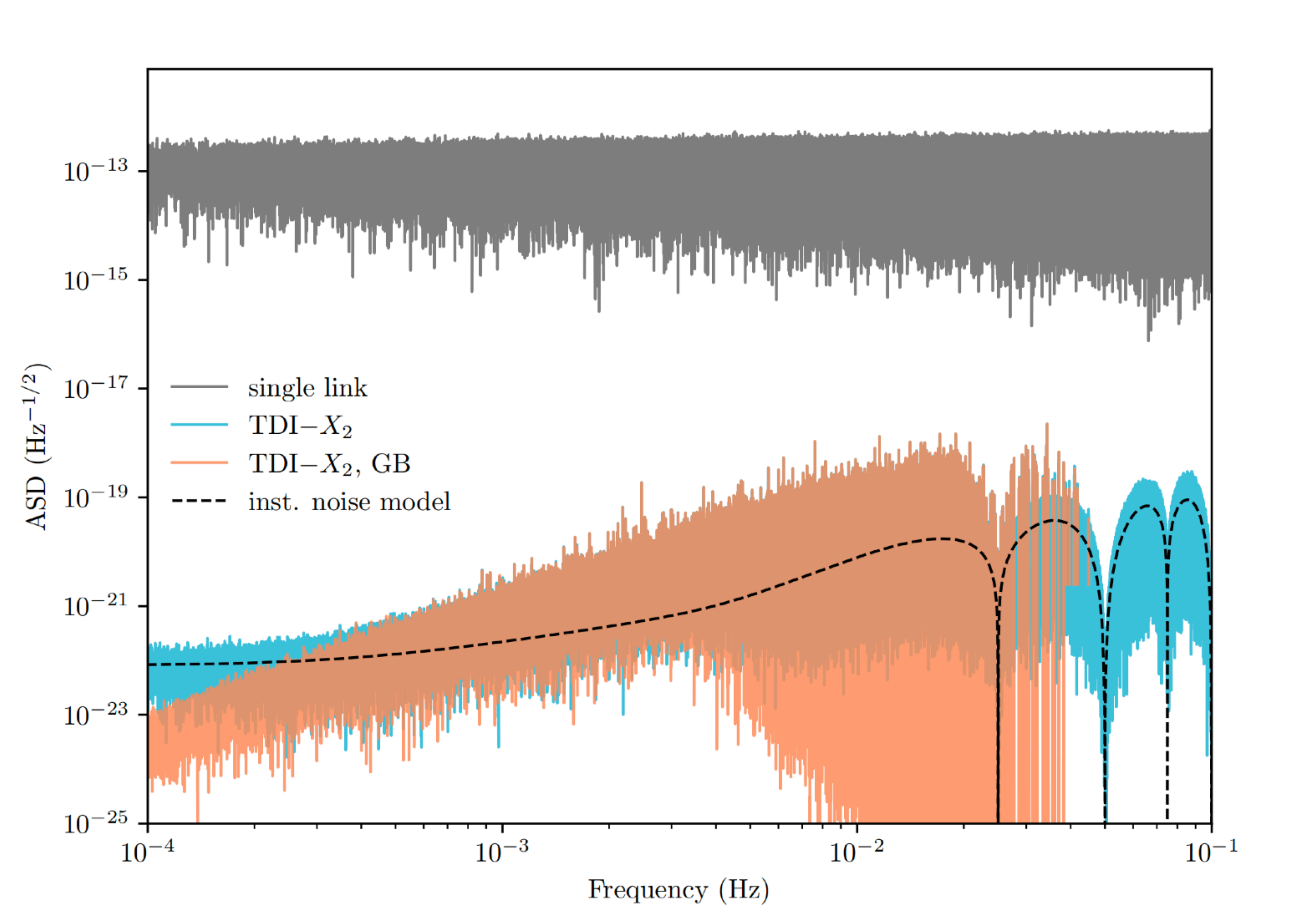
Data in frequency domain (ASD)



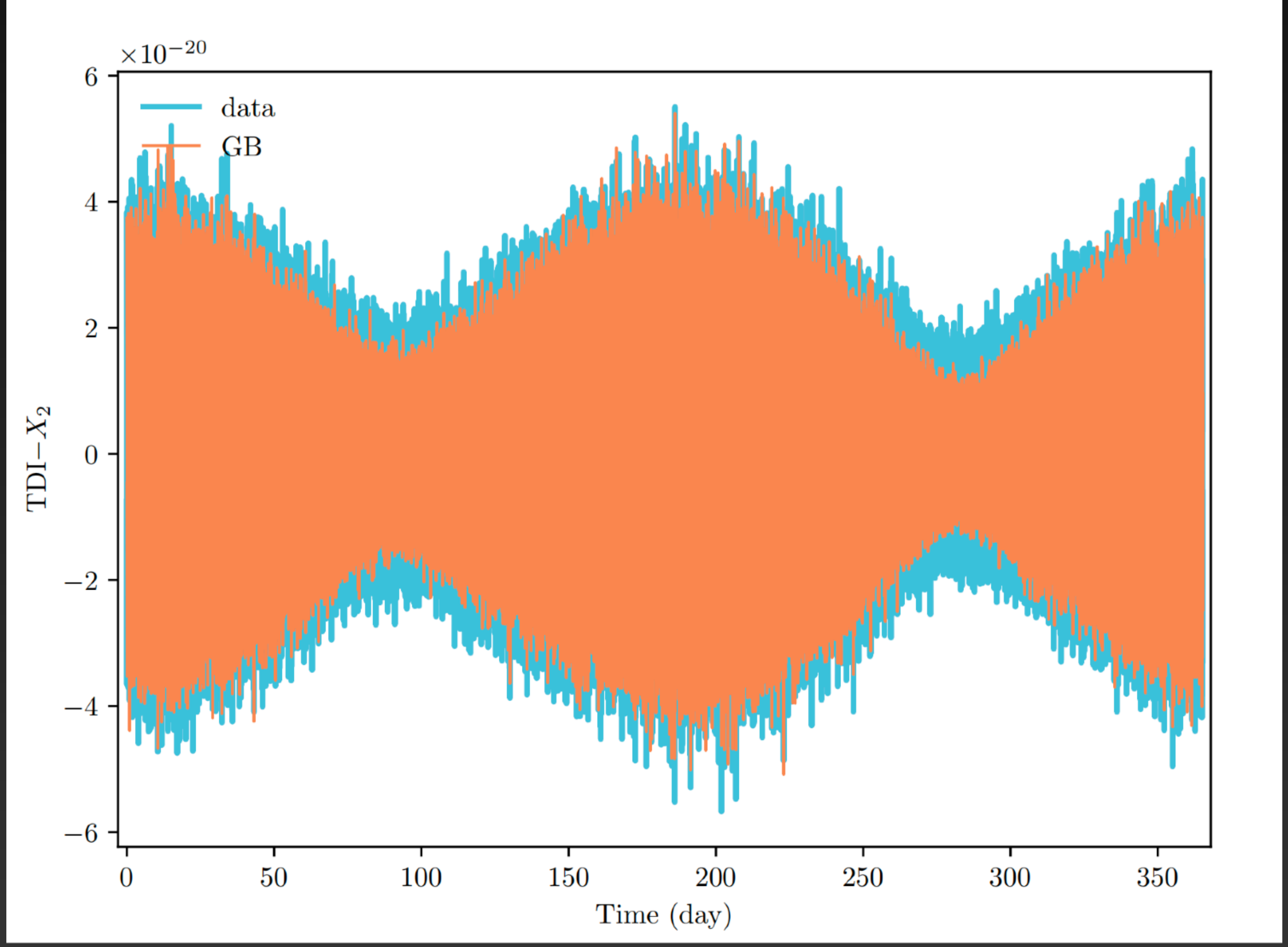
Data in time domain (downsampled to 0.1 Hz)

Challenge 2: The "Single Source" Datasets

Target	Search of overlapping GBs
Dataset 2.1	4.5e7 GBs + inst. noise



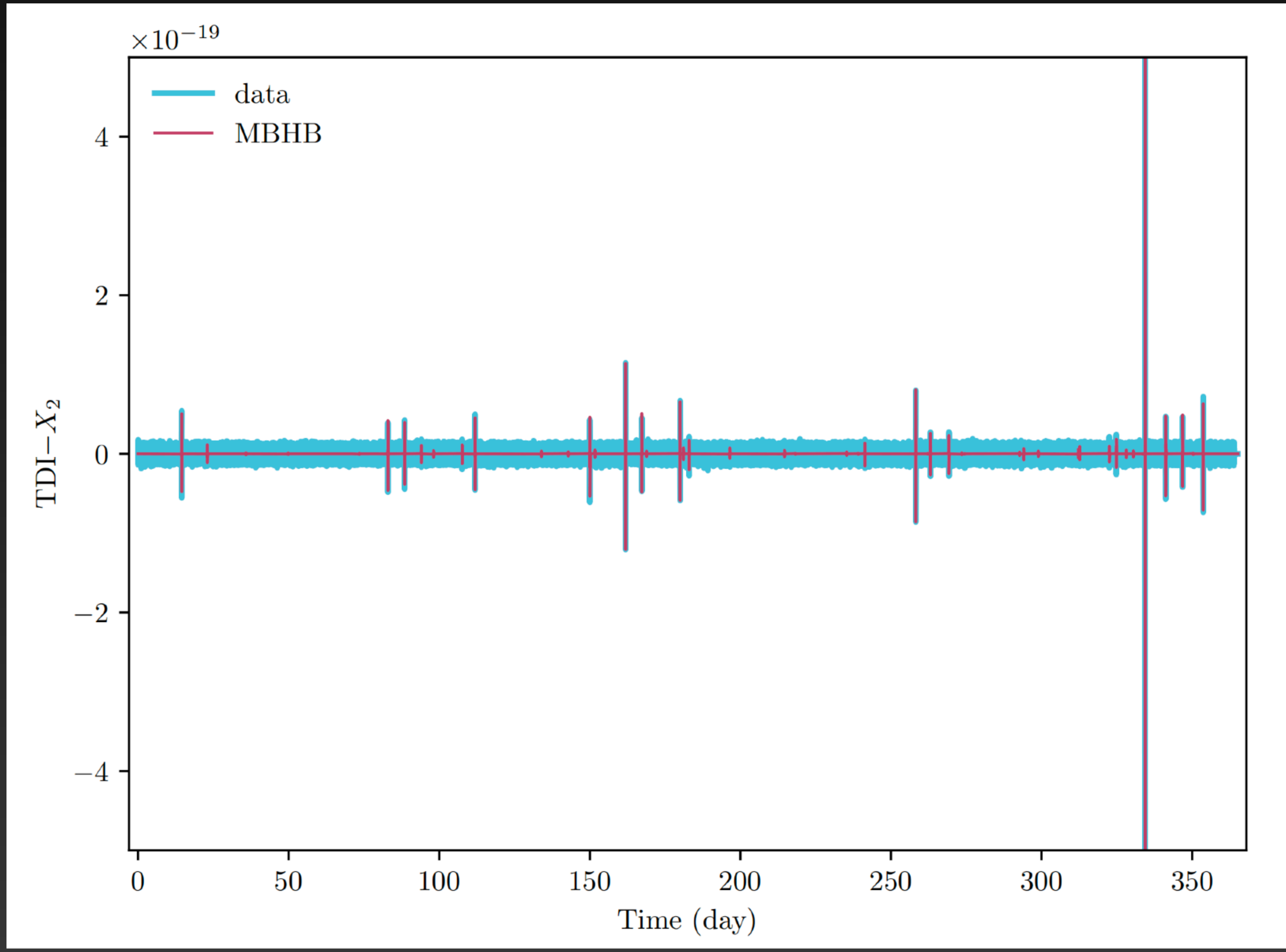
Data in frequency domain (ASD)



Data in time domain (downsampled to 0.1 Hz)

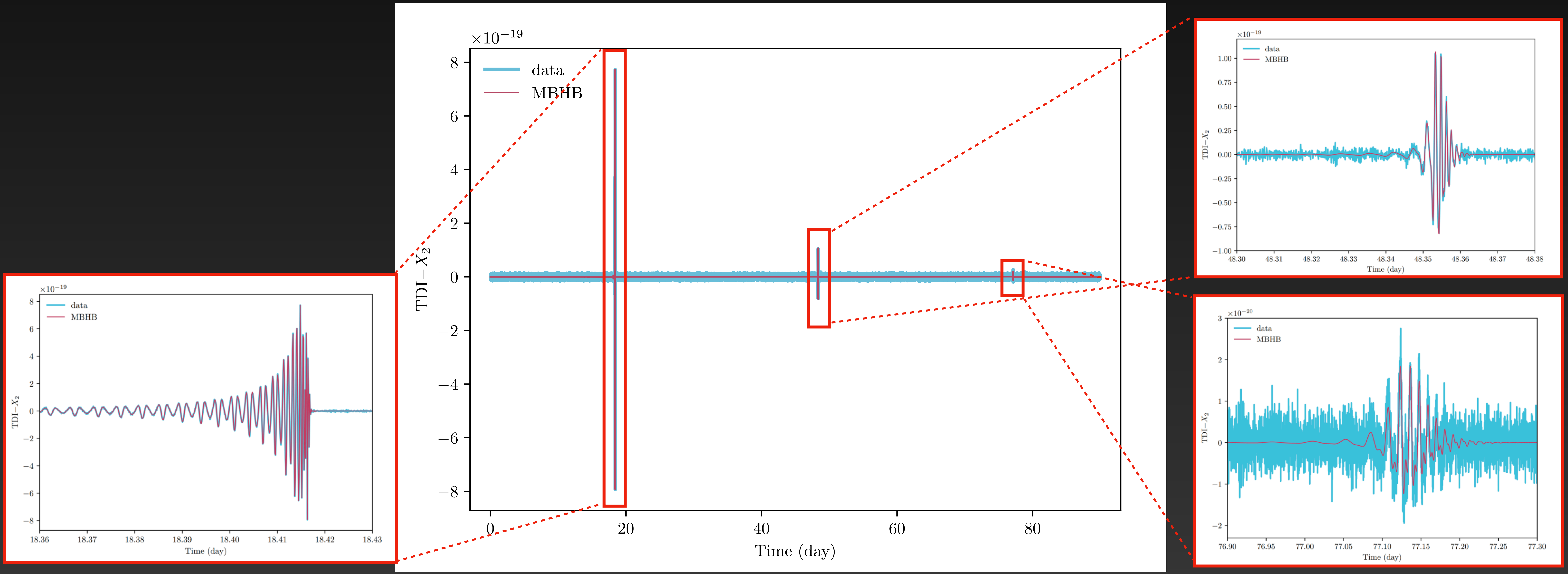
Challenge 2: The “Single Source” Datasets

Target	Estimation of more overlapping MBHBs
Dataset 2.2	50 MBHBs (PhenomT) + inst. noise



Challenge 2: The “Single Source” Datasets

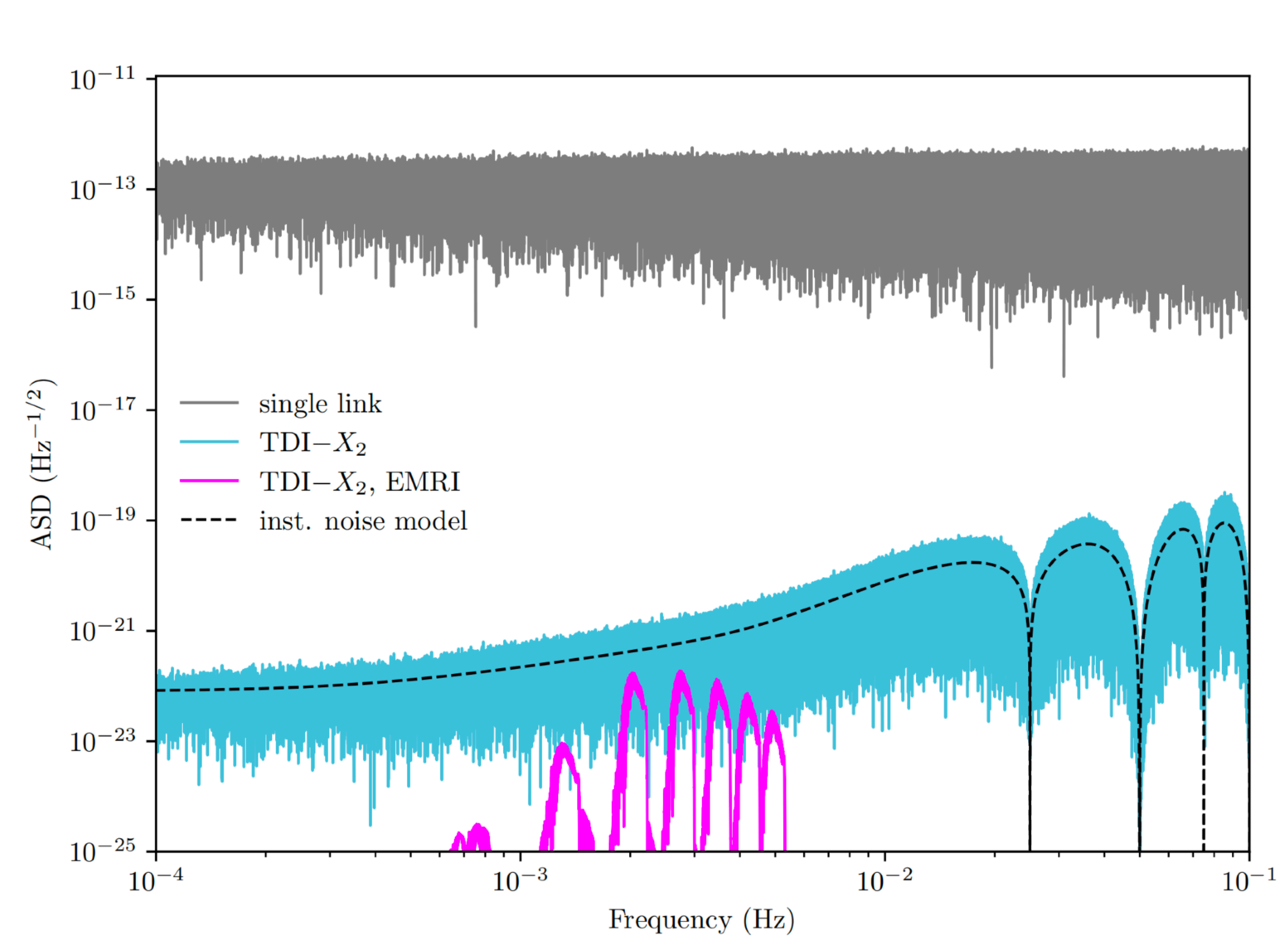
Target	Estimation of MBHBs with complex waveforms
Dataset 2.3	3 MBHBs (SEOBNRv5EHM)



The eccentric high-harmonic MBHB dataset

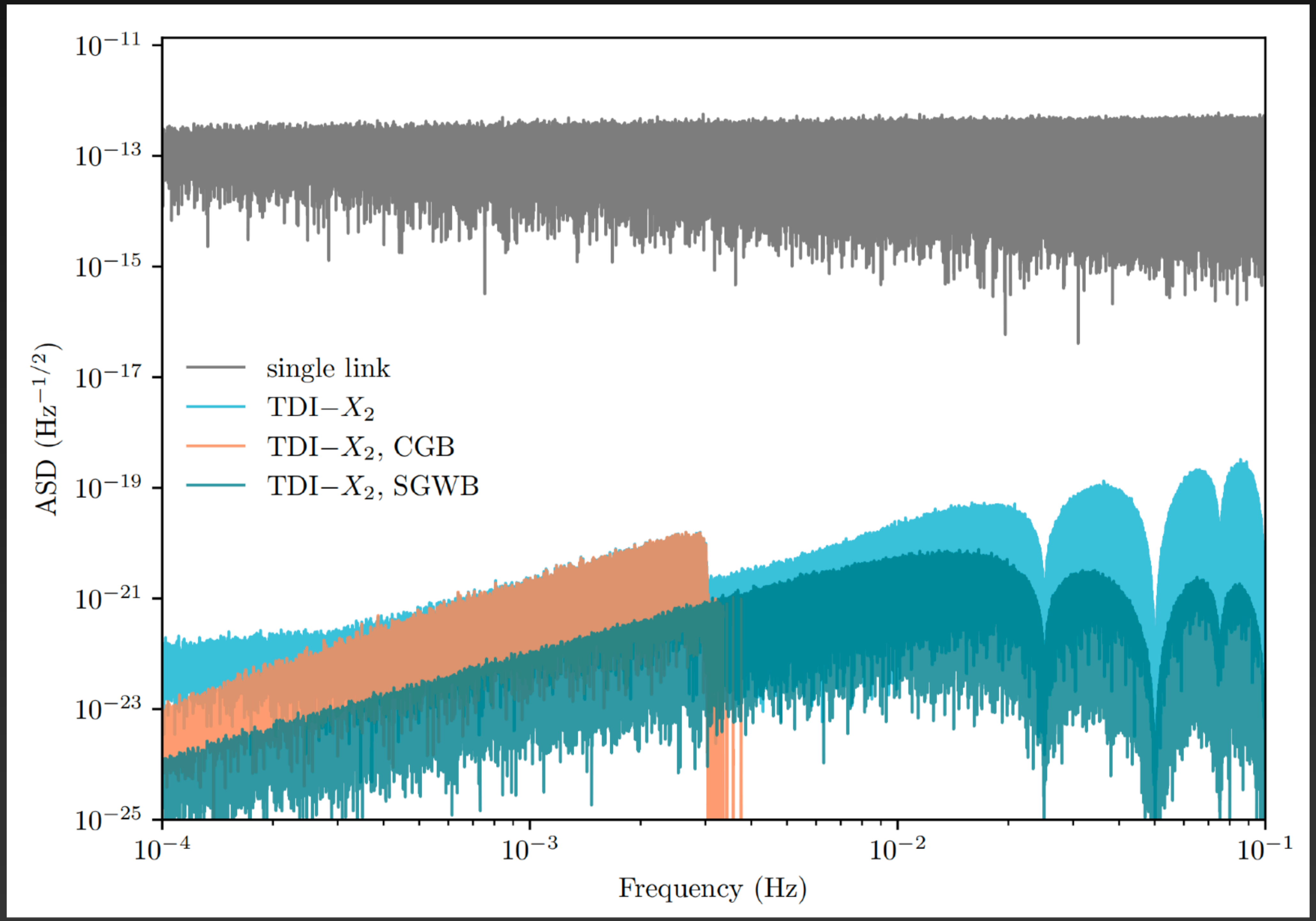
Challenge 2: The "Single Source" Datasets

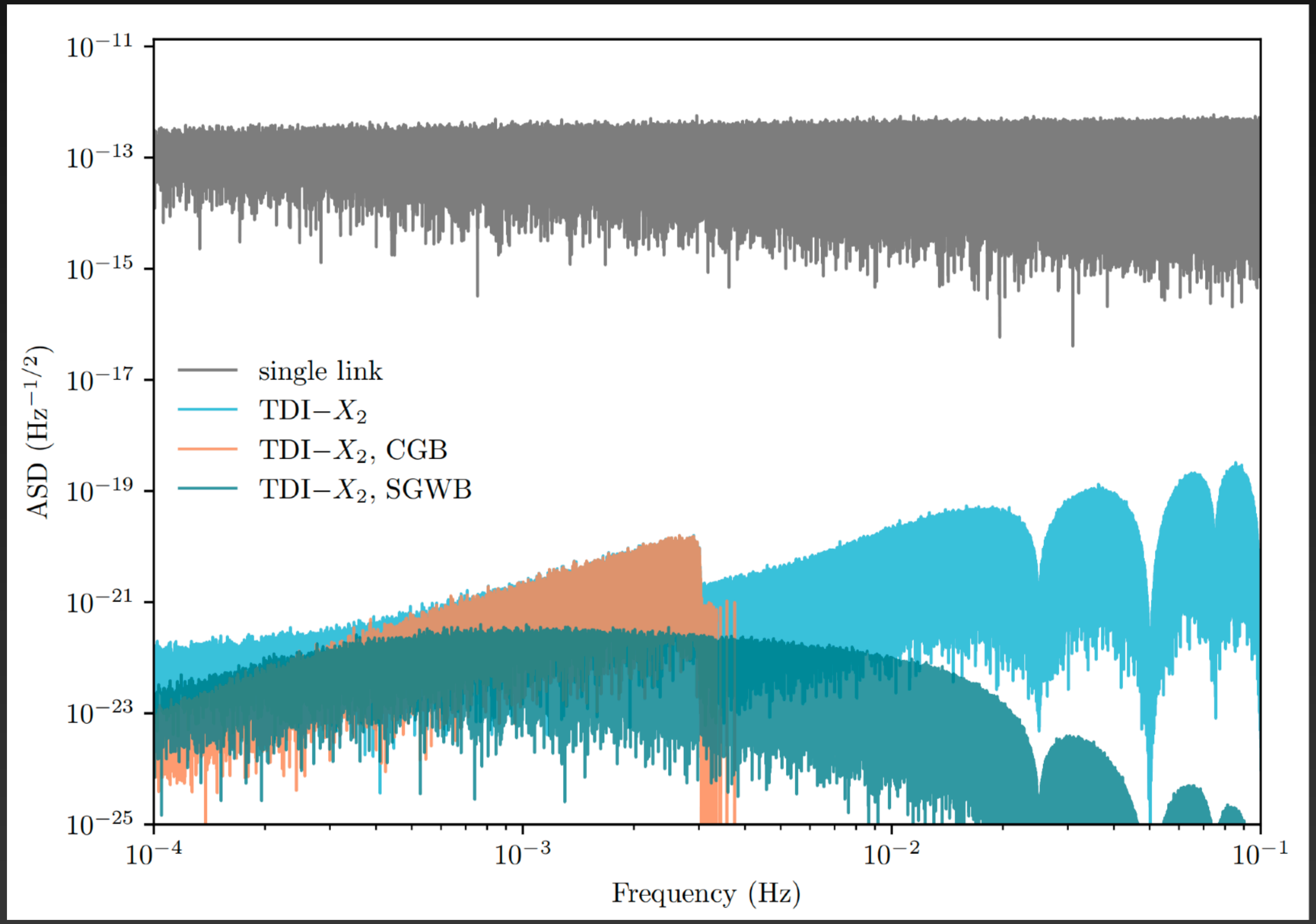
Target	Estimation of EMRIs
Dataset 2.4/2.5/2.6/2.7	1 EMRI (AK/Kerr/Bumpy/b-EMRI)



Challenge 2: The "Single Source" Datasets

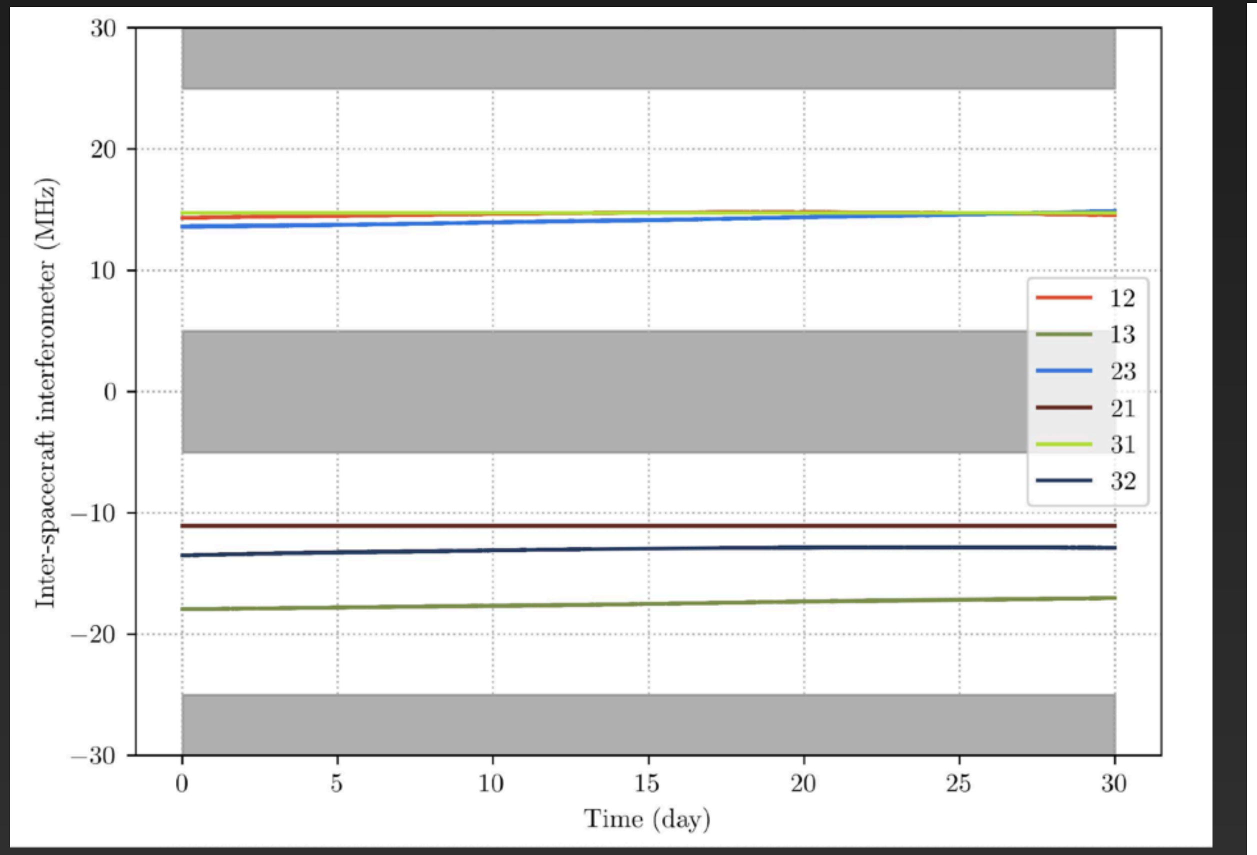
Target	Estimation of SGWB under the confusion of foreground & inst. noise
Dataset 2.8/2.9	SGWB (different shapes)+GB foreground (no bright sources)+inst. noise

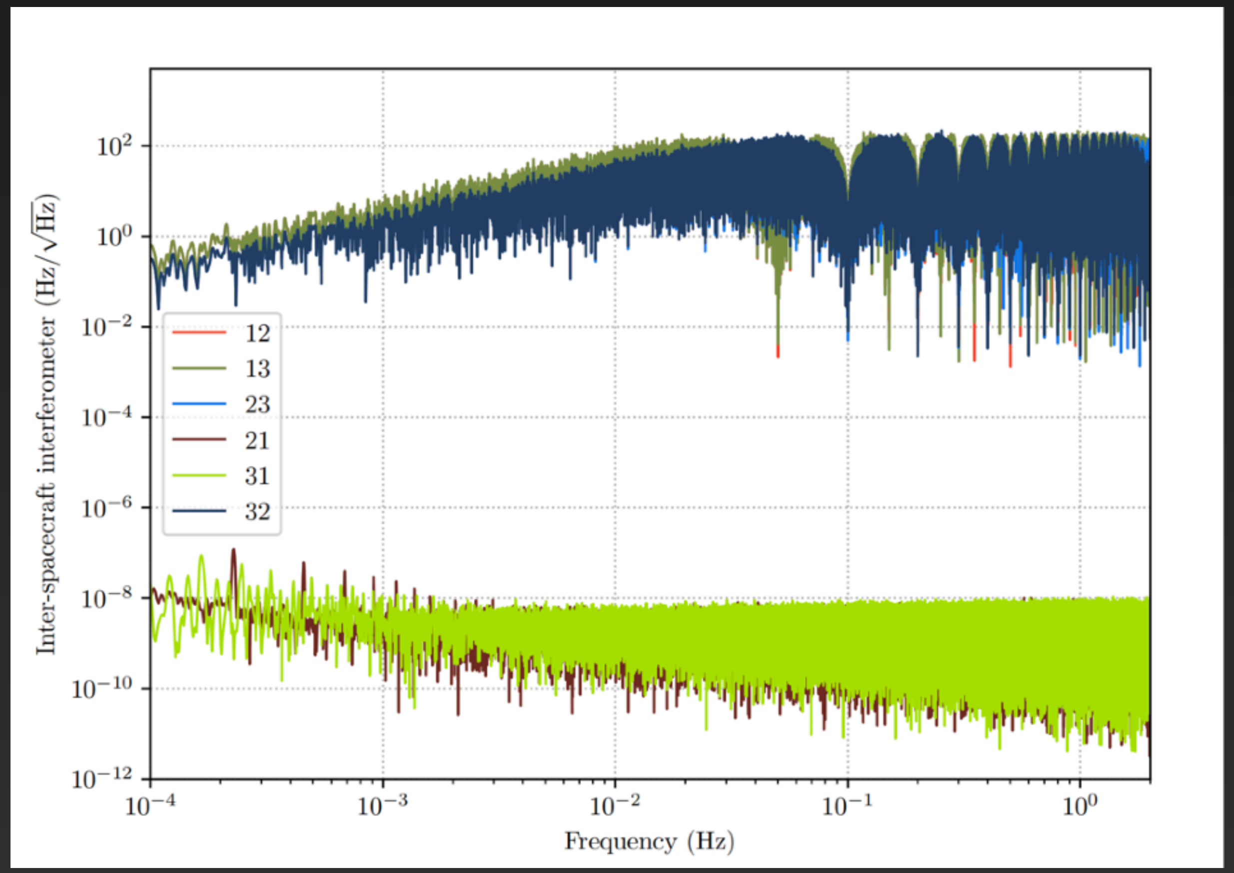




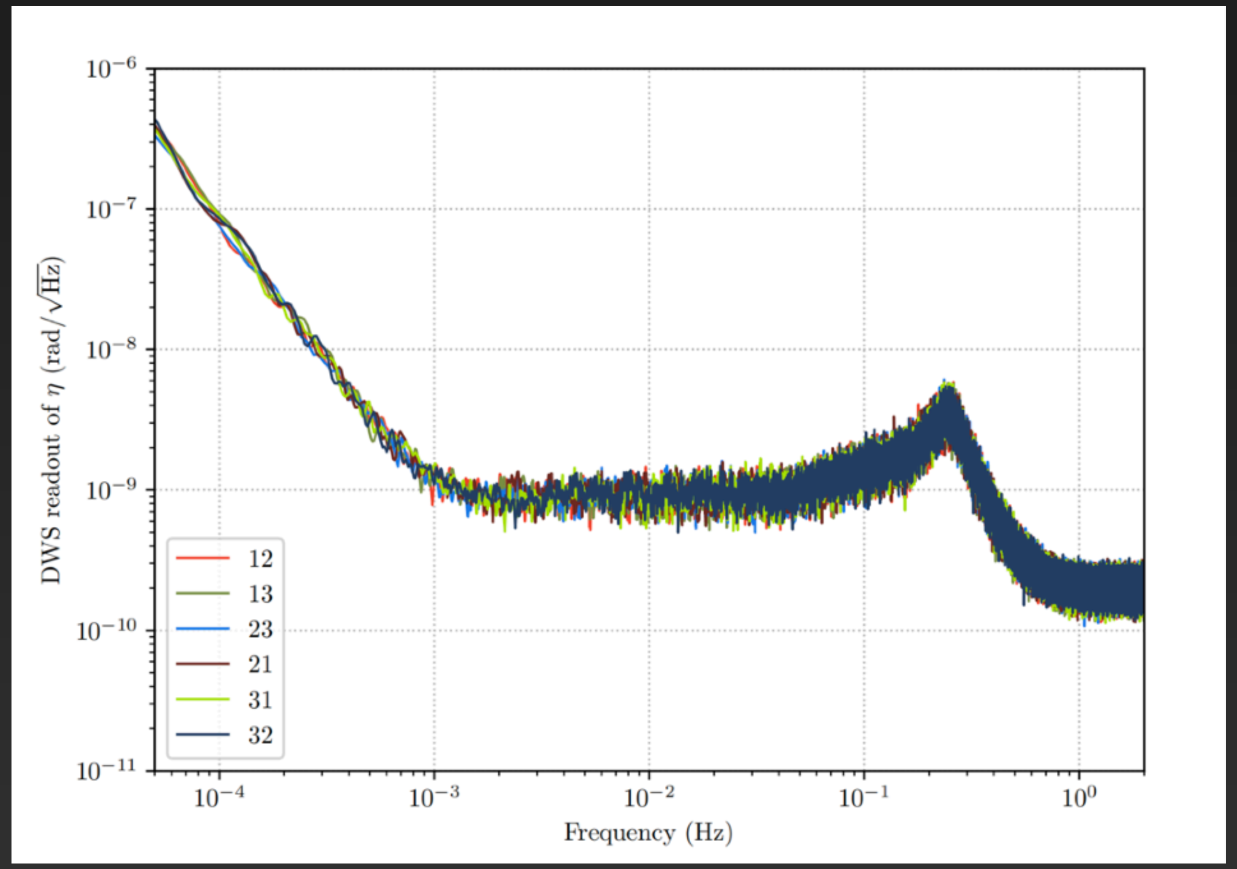
Challenge 3: The “End-to-End” Dataset

Target	Address the coupling problems of pre-processing and scientific analysis & build an end-to-end pipeline
Dataset 3.1	1 MBHB, 21 GBs, all listed inst. noise, clock deviations
Data form	Raw interferometric measurements, differential wavefront sensors (DWSs), pseudo ranges, orbit determination, measured clock deviations

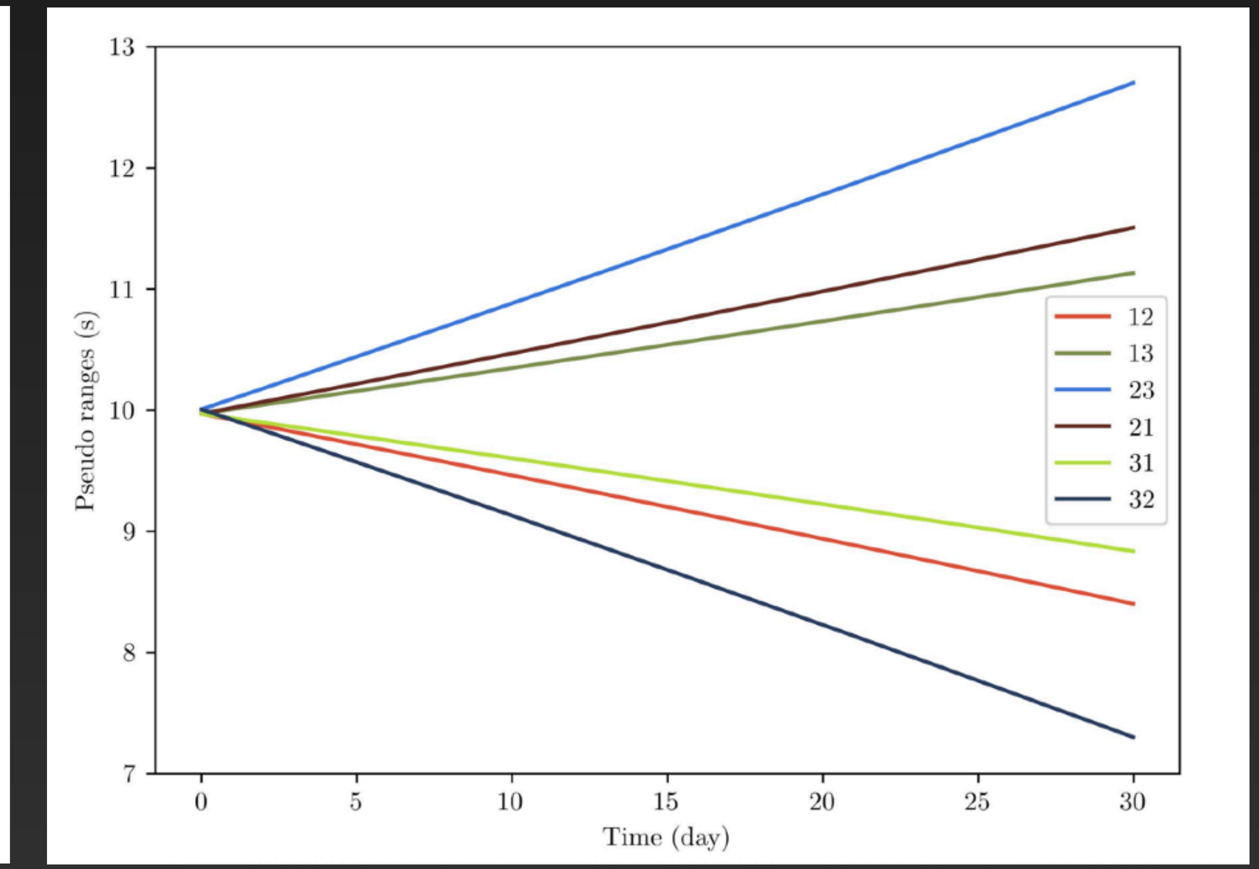




Raw interferometric data in time / frequency domain



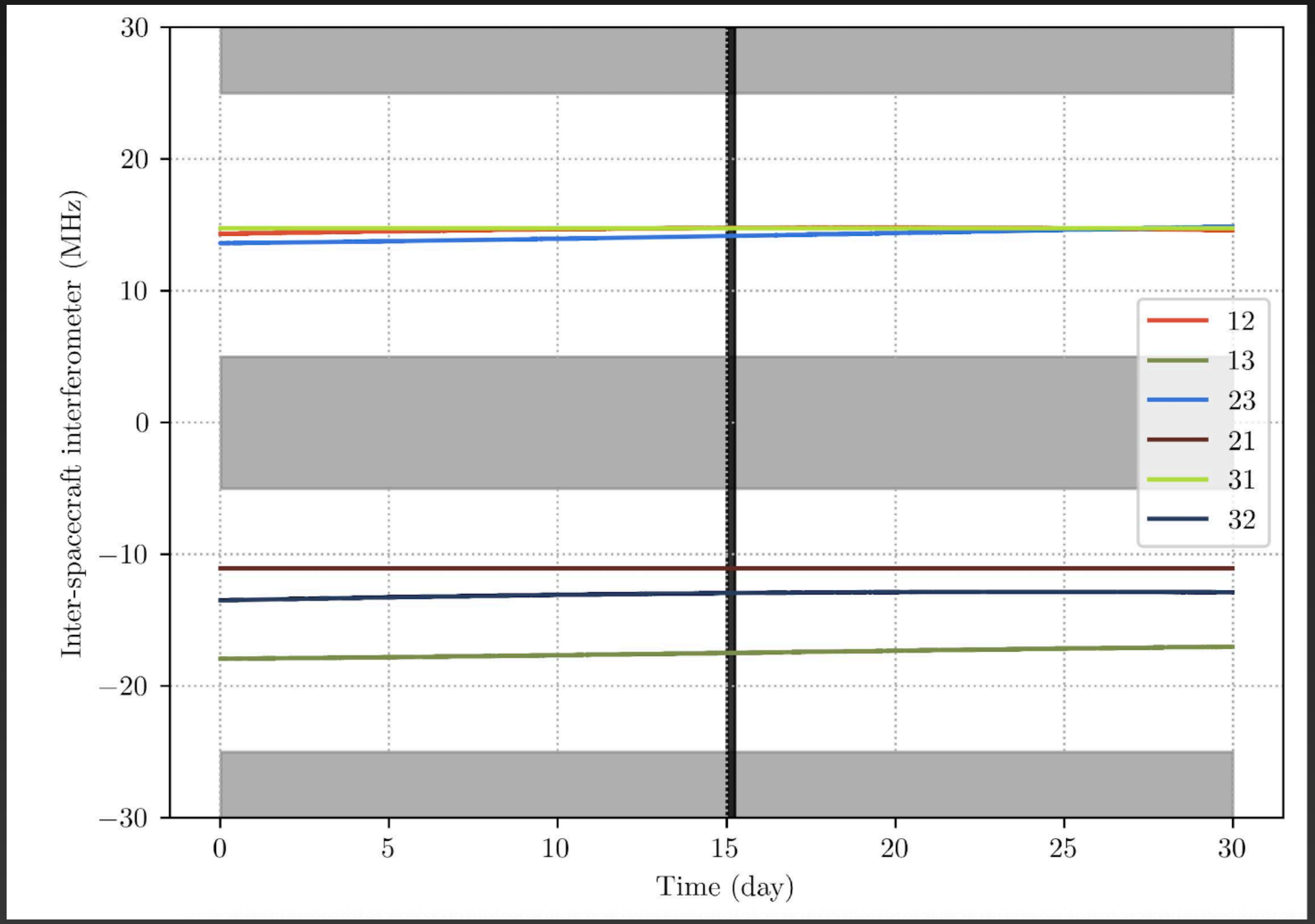
DWS in frequency domain



Pseudo ranges

Challenge 4: The "Up-to-Date" Dataset

Target	Showcase the progress in Taiji simulation & the up-to-date knowledge on data features
Dataset 4.1	Signals + inst. noise + gaps + glitches
Data form	Raw interferometric measurements, differential wavefront sensors (DWSs), pseudo ranges, orbit determination, measured clock deviations



1. Basic Concepts
2. Mock Data Pipeline
3. Challenges
4. The TDC Datasets
5. **Todos**
6. Triangle
7. Tests and Verifications
8. Take-away messages

Todos: other data analysis challenge & GW science challenges

Astrophysics	The production and growth of black hole seeds The co-evolution of massive black holes and their host galaxies The existence and abundance of intermediate-mass black holes The environments of massive black holes at the galaxy centers The astrophysics of stellar-mass black holes binaries The astrophysics of Milky Way galaxy ……
Cosmology	The origin and distribution of dark matter The Hubble constant and other cosmological parameters ……
Fundamental physics	The speed of gravitational waves The strong-field dynamics of space time The no-hair theorem The products of black hole mergers The TeV high-energy physics in the very early universe The cosmic strings ……

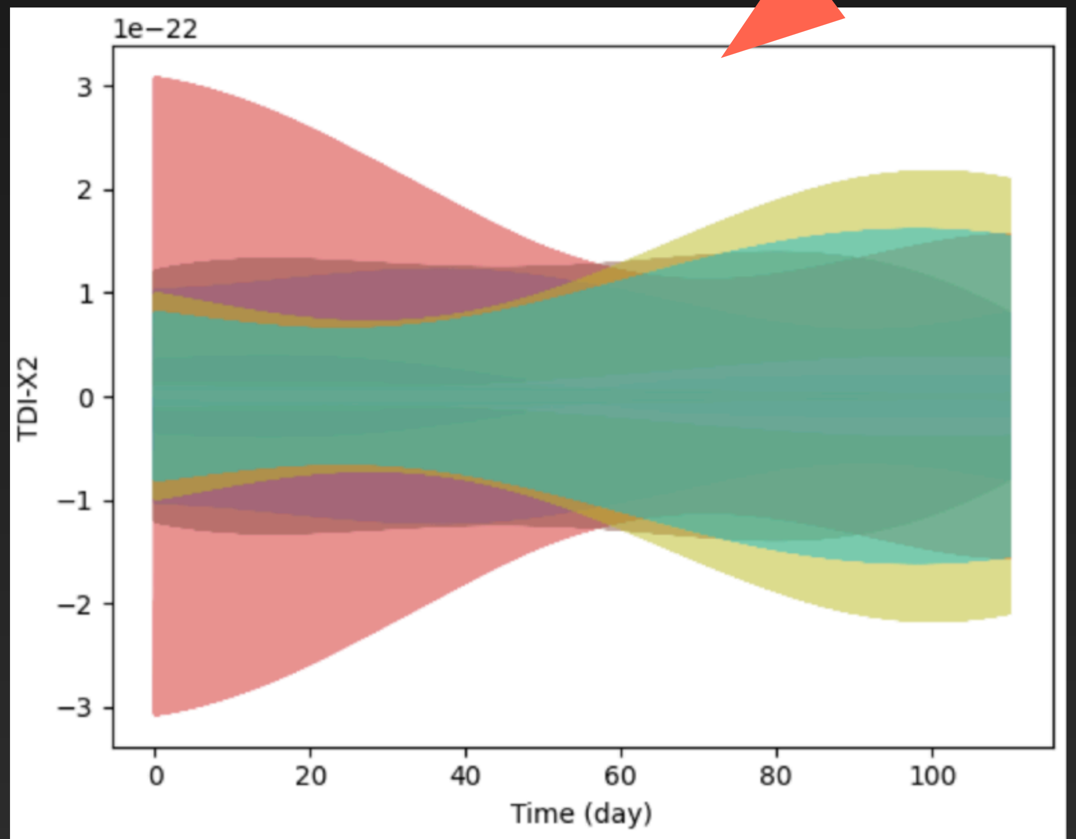
1. Basic Concepts
2. Mock Data Pipeline
3. Challenges
4. The TDC Datasets
5. Todos
6. **Triangle**
7. Tests and Verifications
8. Take-away messages

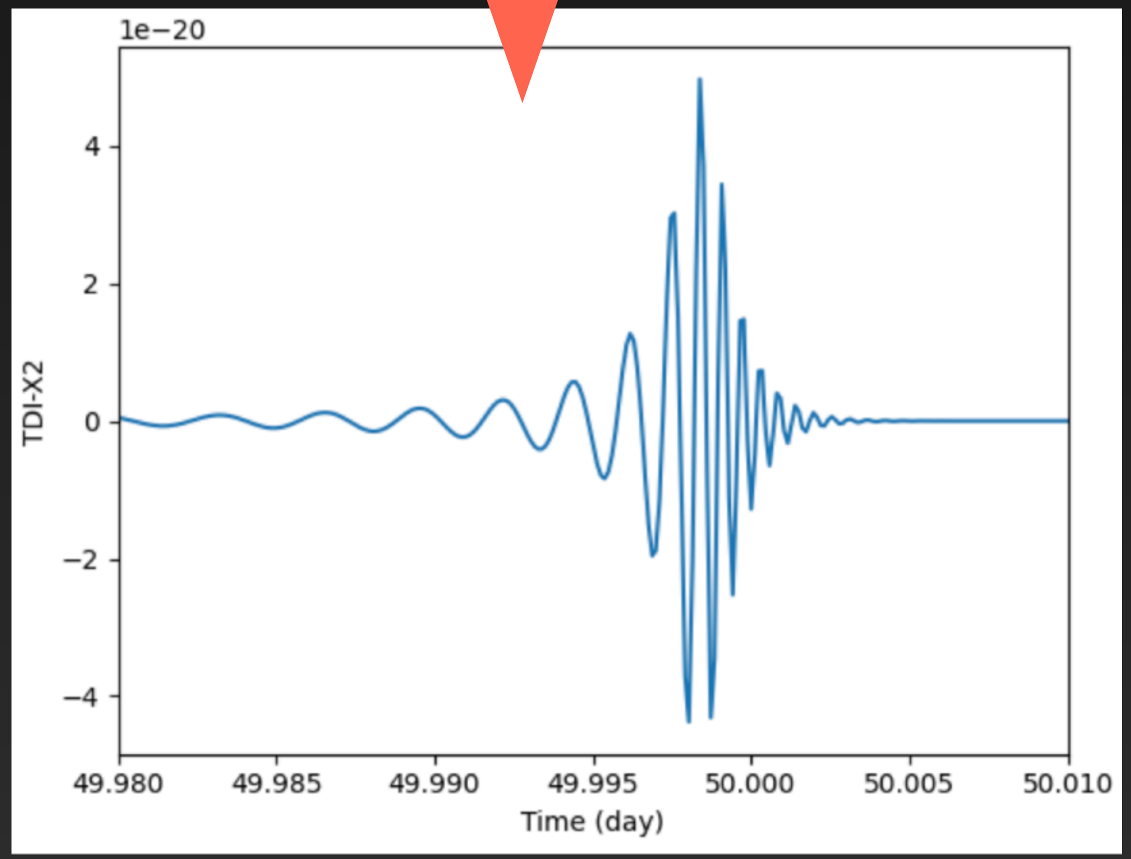
Do the to-dos with Triangle

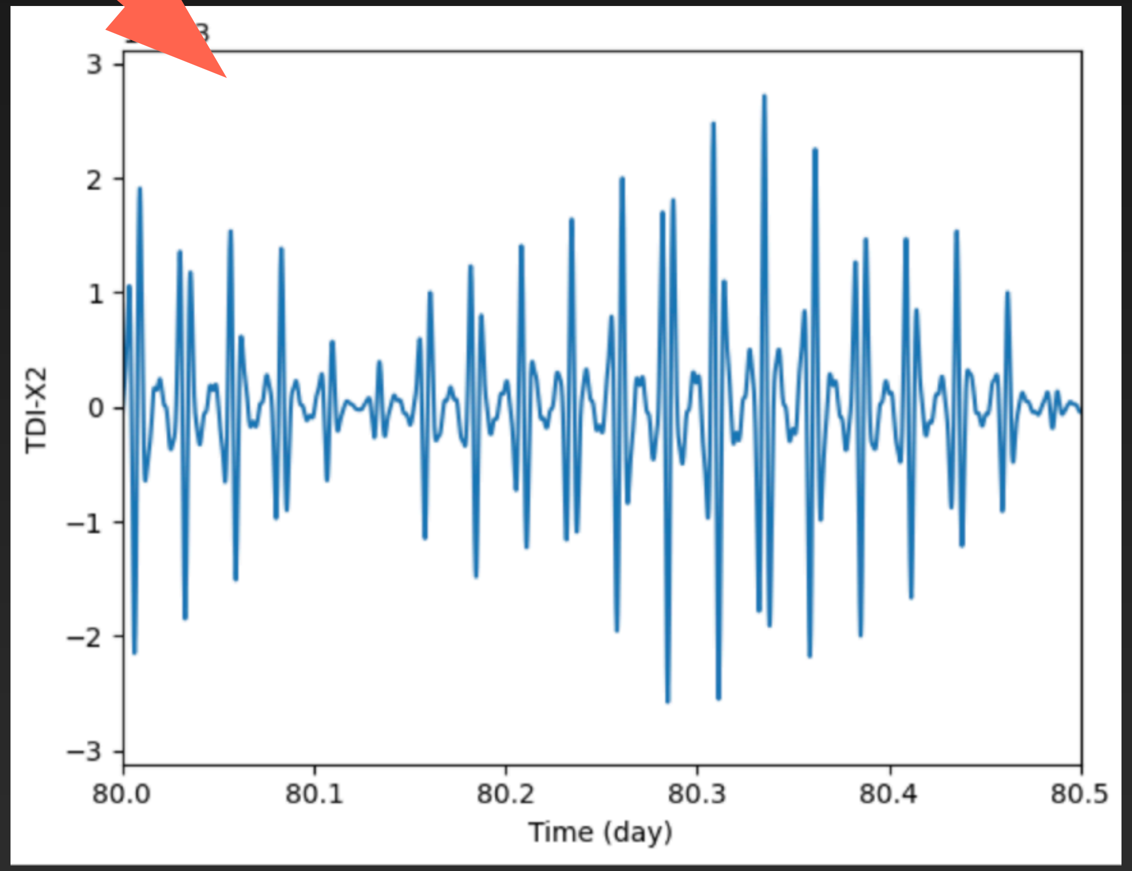
GW signal injection

```python
# Create a generator for the unprojected waveforms
gb_waveform_generator = GB_Injection(use_gpu=False)

# Create a generator for the TDI responses. The initialization takes a while, but it only needs to be run once.
# After the first call, the calculator can be run repeatedly with different source parameters.
# Interpolation is used dfferently from the Interferometer and TDI modules. For GBs (with f0 < 2e-2 Hz), linear
# While we recommand linear_interp = False for more complex waveforms such as MBHBs and EMRIs.
gb_response_generator = GeneralTDIResponse(
    orbit=orbit,
    Pstring=X2_string,
    tcb_times=tcb_times,
    use_gpu=False,
    drop_points=int(1000/dt),
    linear_interp=True,
    )
```







Custom noise simulation

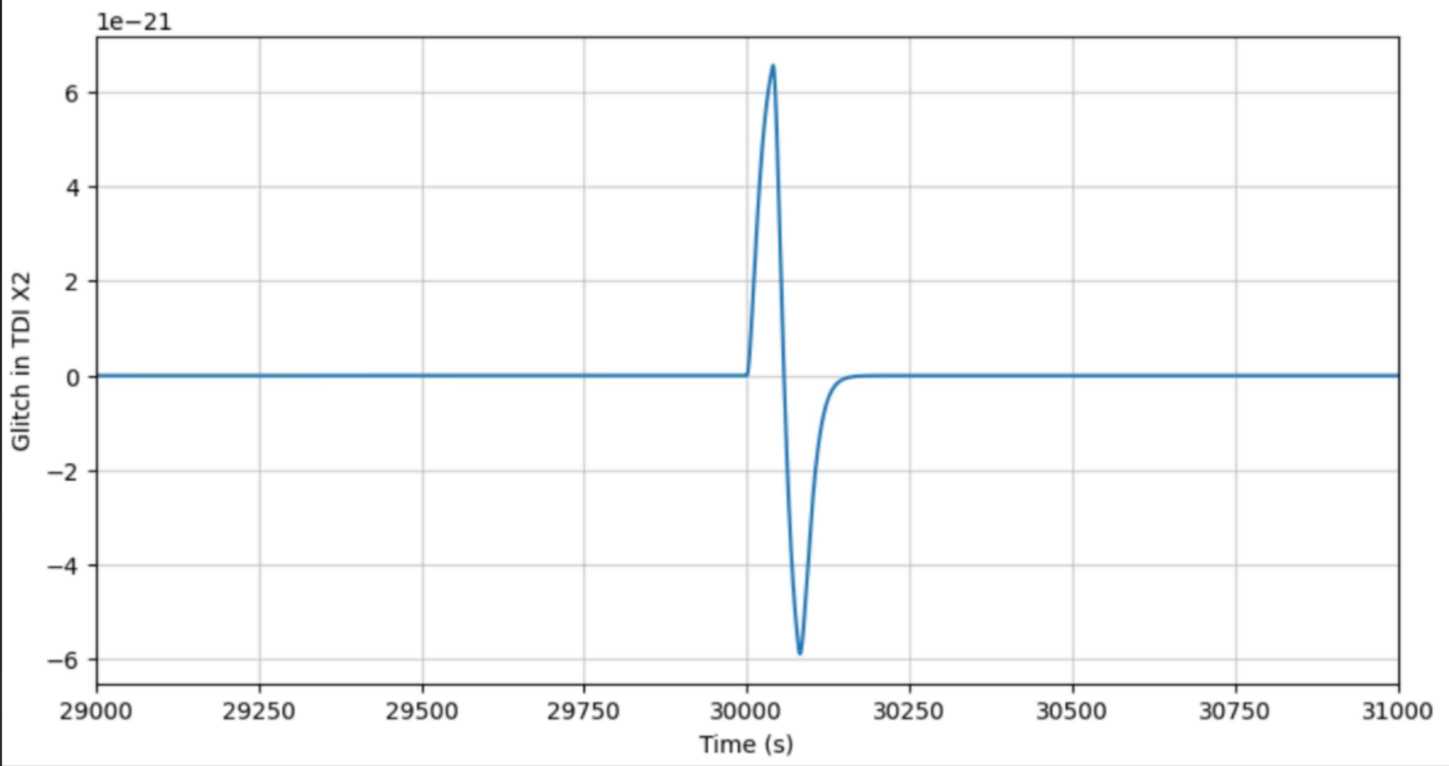
```python
ifo = Interferometers(
    size=size,
    fsample=fsample,
    time_frame='ClockTime',
    acc_noise=True,
    ro_noise=True,
    laser_noise=True,
    clock_noise=True,
    ob_noise=True,
    bl_noise=False,
    op_noise=False,
    ranging_noise=False,
    modulation_noise=False,
    orbit_class=orbit,
    offset_class=offset,
    noise_class=noise,
    gw_class=None,
    fplan=fplan,
    order=interp_order,
    pool=pool,
    clean_memory=True,
    )
```

Arbitrary TDI combination

$$\mathrm{TDI} = \sum_{ij \in \mathrm{MOSAs}} \mathrm{P}_{ij} \eta_{ij}$$

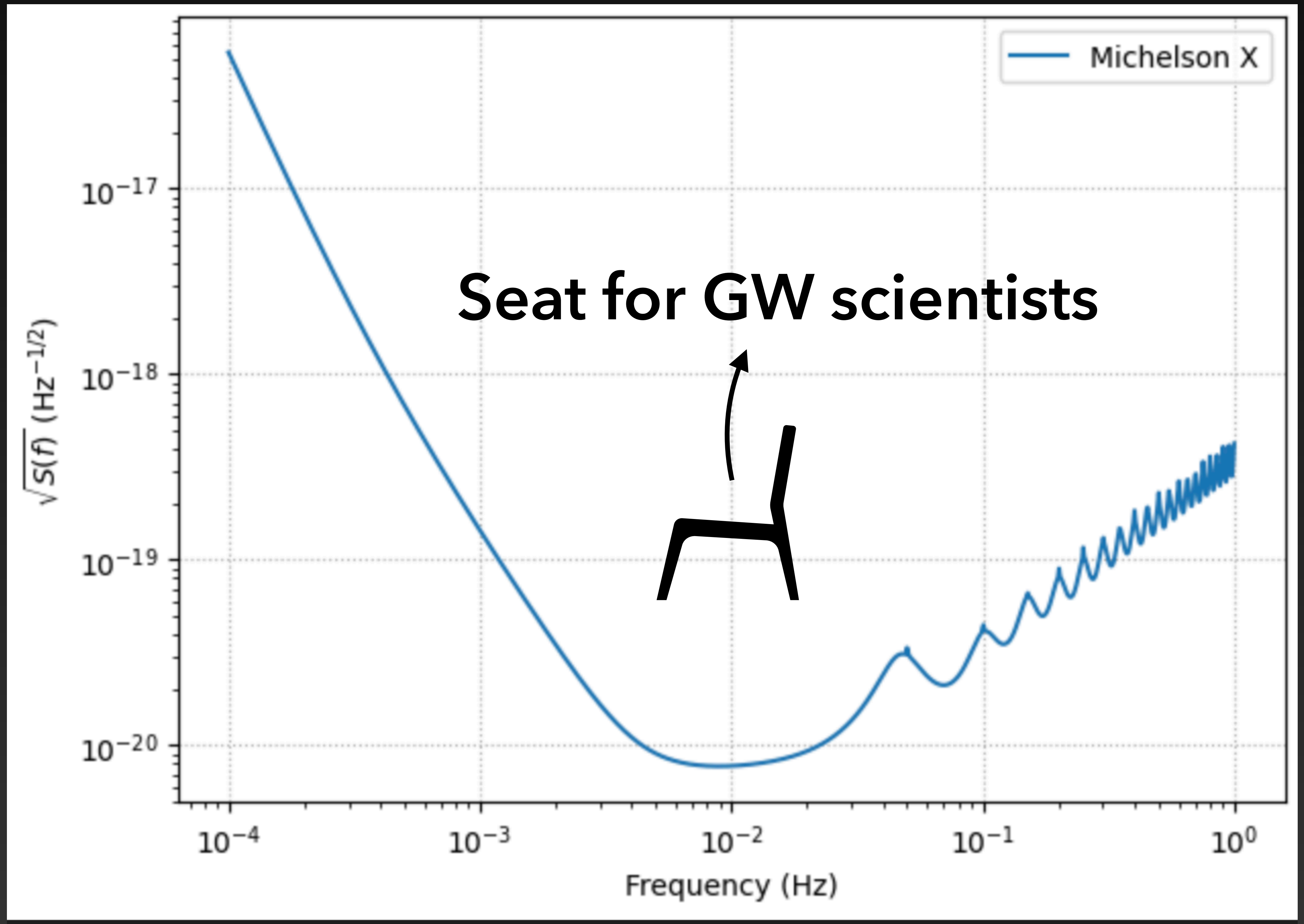
$$\begin{aligned}
\mathbf{P}_{12} &= 1 - \mathrm{D}_{131} - \mathrm{D}_{13121} + \mathrm{D}_{1213131}, \\
\mathbf{P}_{23} &= 0, \\
\mathbf{P}_{31} &= -\mathrm{D}_{13} + \mathrm{D}_{1213} + \mathrm{D}_{121313} - \mathrm{D}_{13121213}, \\
\mathbf{P}_{21} &= \mathrm{D}_{12} - \mathrm{D}_{1312} - \mathrm{D}_{131212} + \mathrm{D}_{12131312}, \\
\mathbf{P}_{32} &= 0, \\
\mathbf{P}_{13} &= -1 + \mathrm{D}_{121} + \mathrm{D}_{12131} - \mathrm{D}_{1312121}.
\end{aligned}$$

Glitch injection



Sensitivity analysis

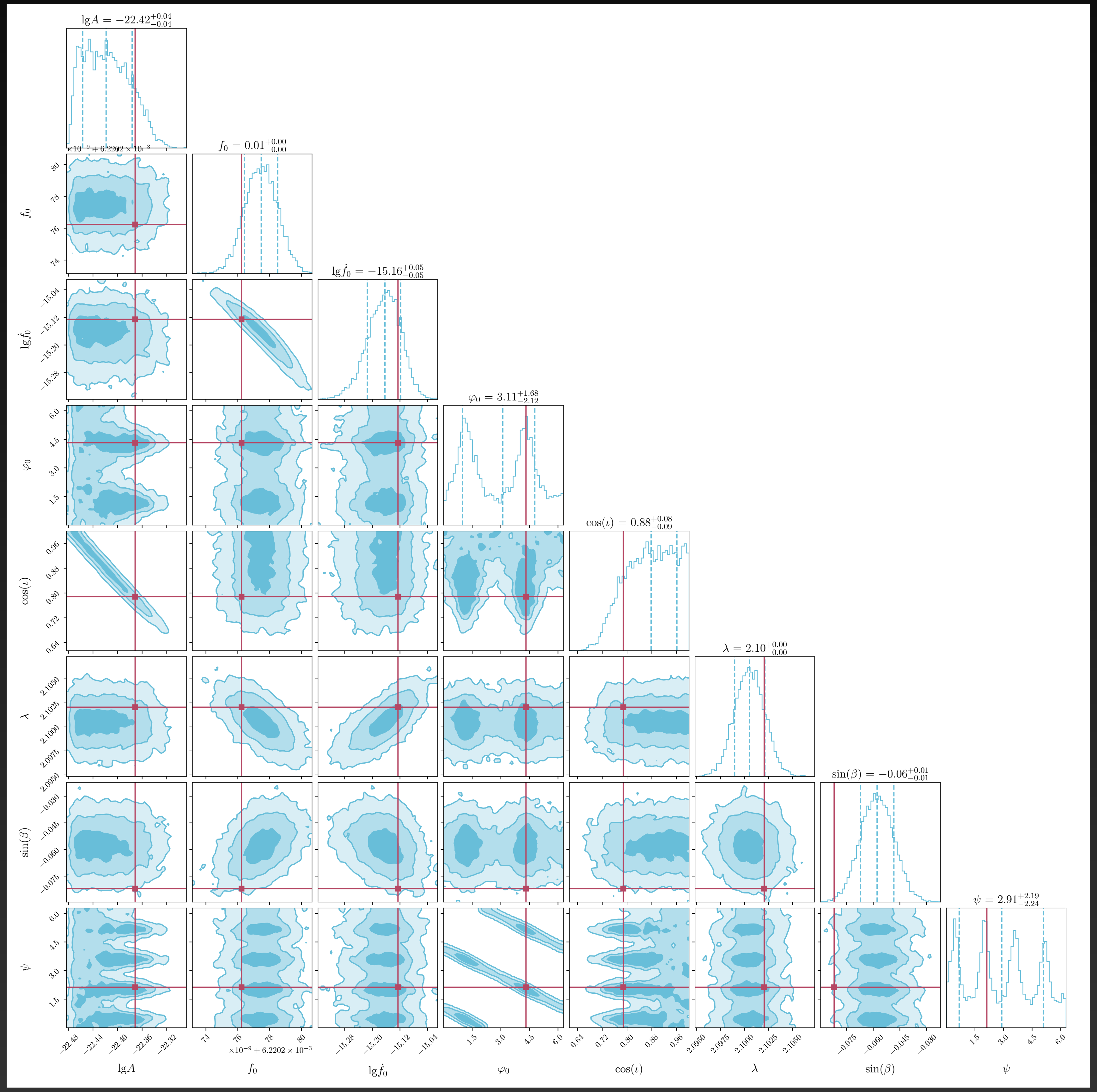
```
sensitivity = TDISensitivity(Ri=Ri_test, nij=nij_test, dij=dij_test, S_OMS=SOMS_nominal, S_ACC=SACC_nominal)
```



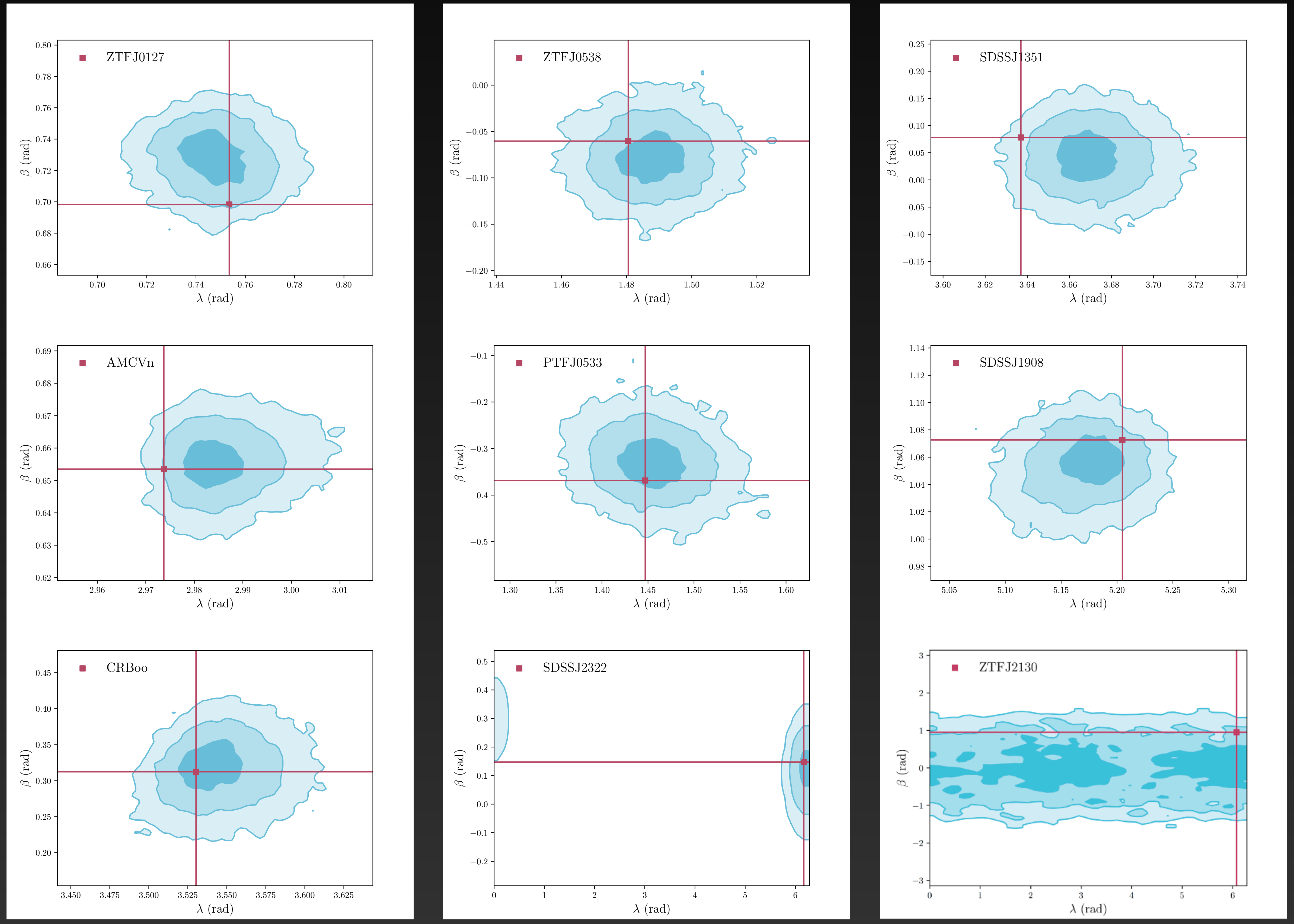
1. Basic Concepts
2. Mock Data Pipeline
3. Challenges
4. The TDC Datasets
5. Todos
6. Triangle
7. **Tests and Verifications**
8. Take-away messages

Test: Parameter Estimation for VGBs

Frequency-domain GB analysis tool: Triangle-GB



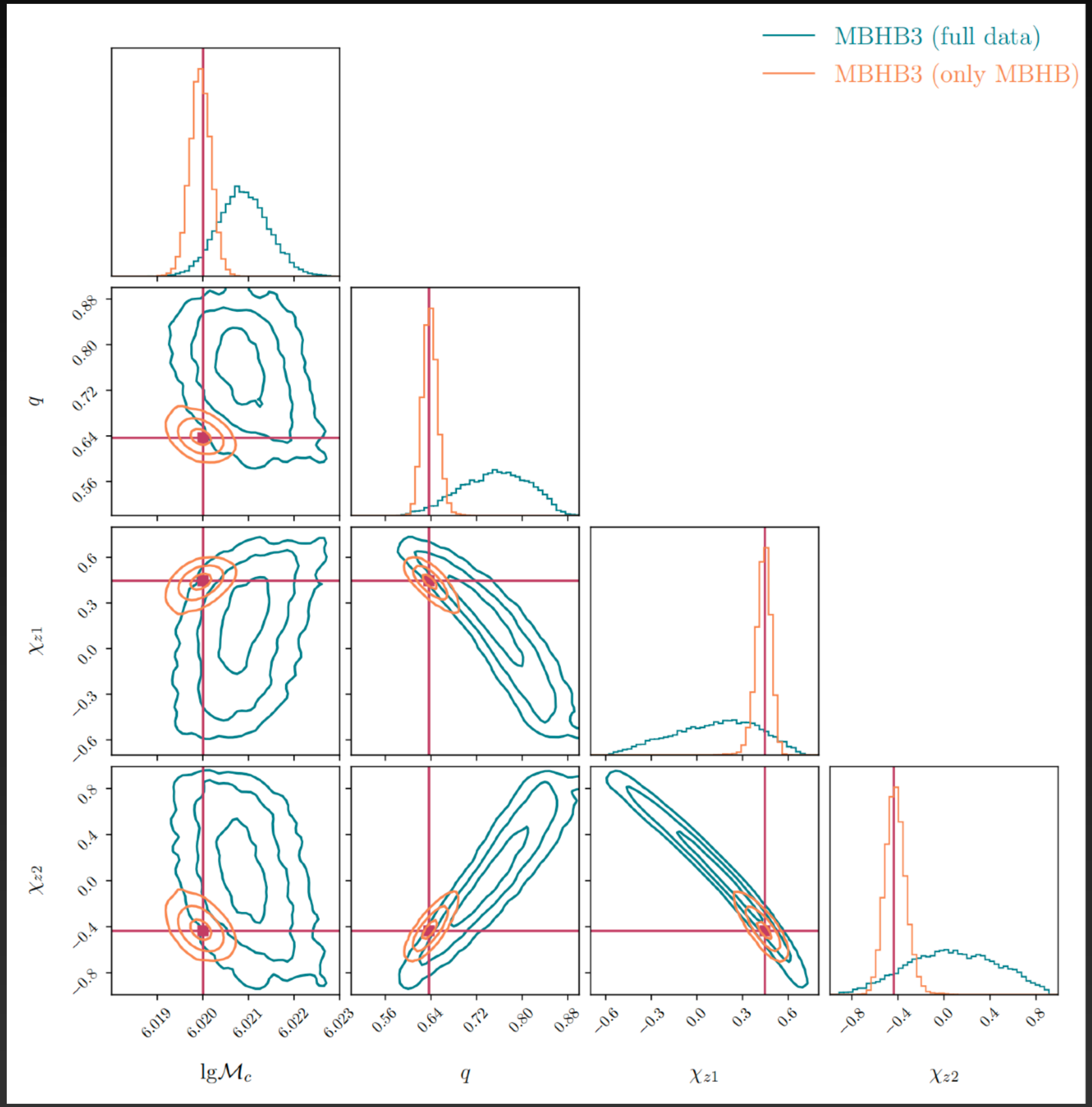
Full posterior of HMCnc



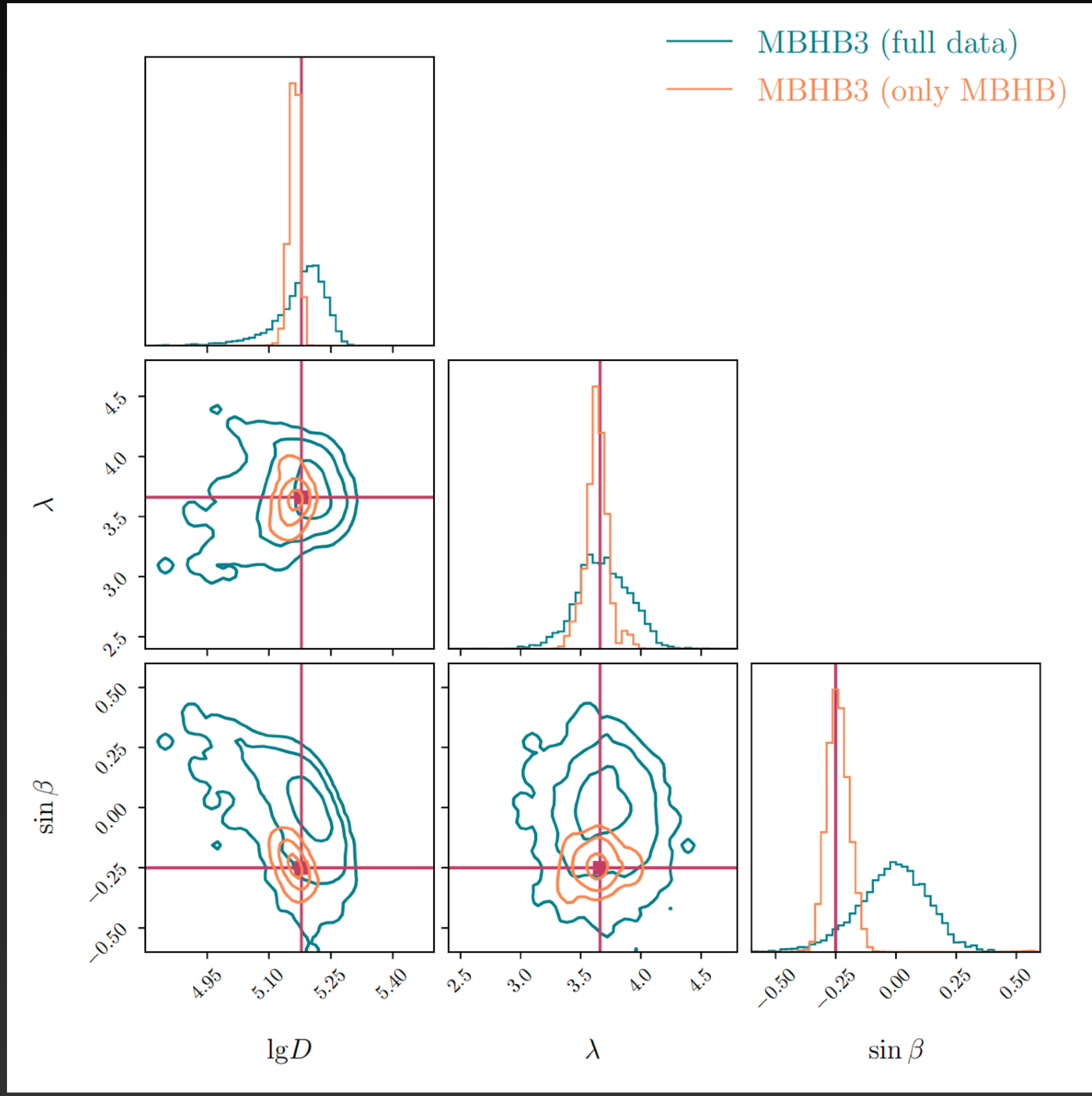
Sky locations of 9 VGBs

Test: Parameter Estimation for MBHBs

Frequency-domain MBHB analysis tool: Triangle-BBH



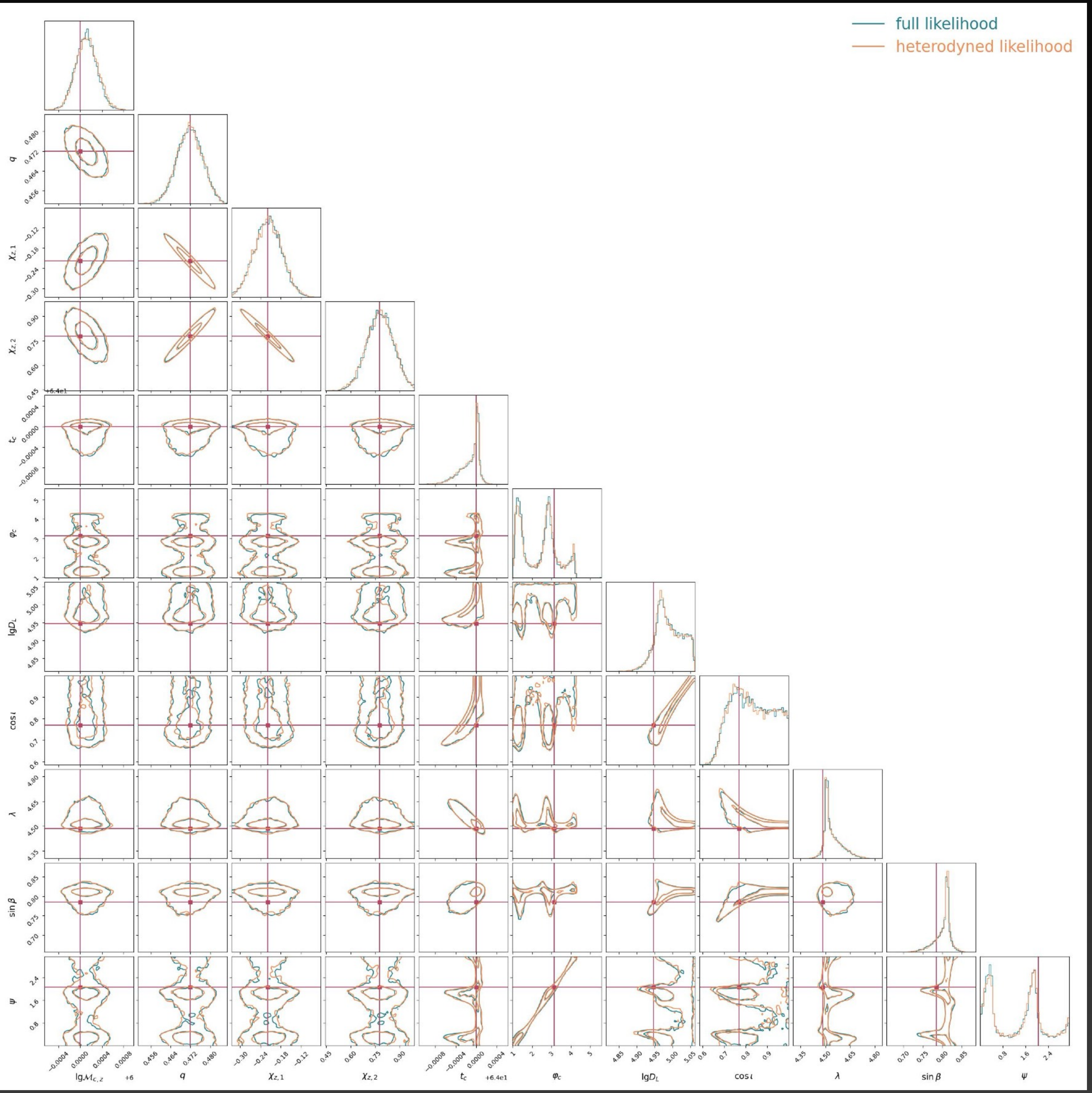
Posterior of intrinsic parameters



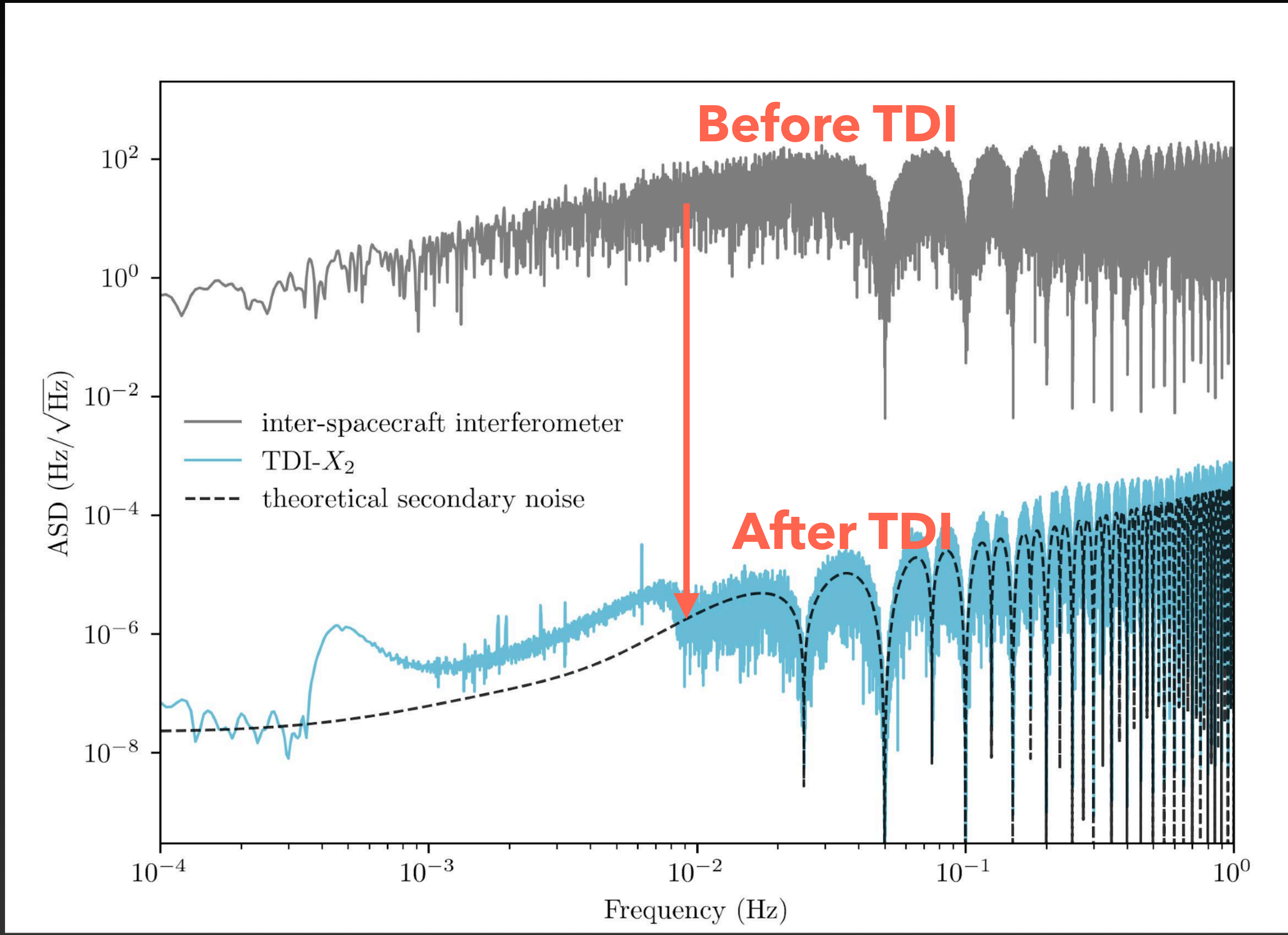
Posterior of 3D location

Test: Parameter Estimation for MBHBs

Full likelihood vs heterodyned likelihood



Test: new TDI topology on raw unsynchronized data



1. Basic Concepts
2. Mock Data Pipeline
3. Challenges
4. The TDC Datasets
5. Todos
6. Triangle
7. Tests and Verifications
8. **Take-away messages**

We have prepared a playground for GW scientists and data analyzers

TDC II & Triangle is available now