The TDC working group

Taiji Data Challenge

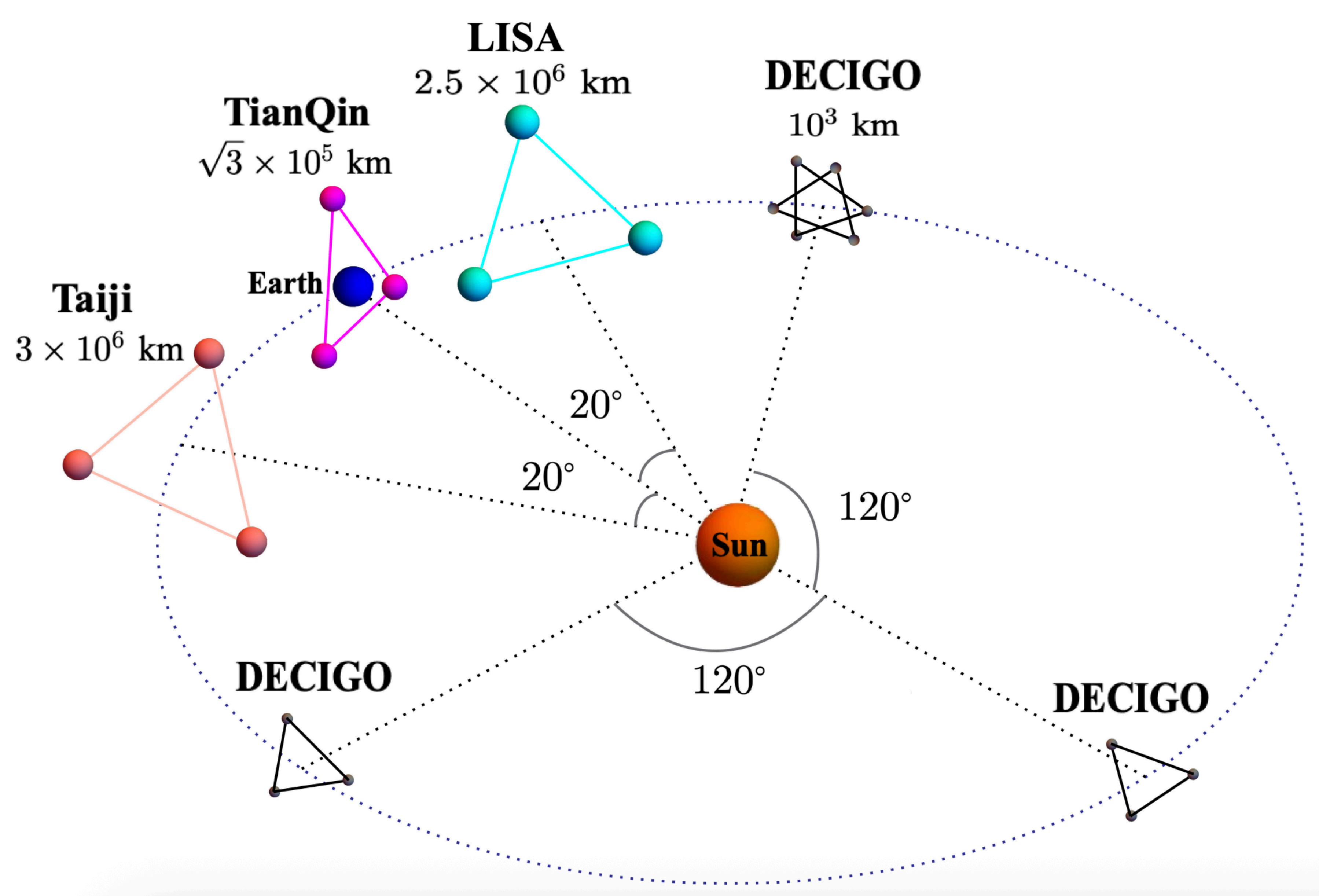
Enhancing Taiji's Estimation on **Galactic Binaries** and **Instrumental Noises** Against Non-Stationaries with Time-Frequency Domain Formalism

PostDoc Minghui Du （杜明辉）

Institute of Mechanics, Chinese Academy of Sciences

Based on **arXiv: 2505.16500** & **arXiv: 2506.10599**

@TAUP2025



[Y. Gong et al, Nature Astronomy 5, 881-889 (2021)]

Taiji is a space-based GW detection project to be launched in the next decade. It basically shares similar mission concepts with LISA and Tianqin, while Taiji and LISA orbit around the sun and Tianqin orbits around the earth.

To whom this talk might be interesting:

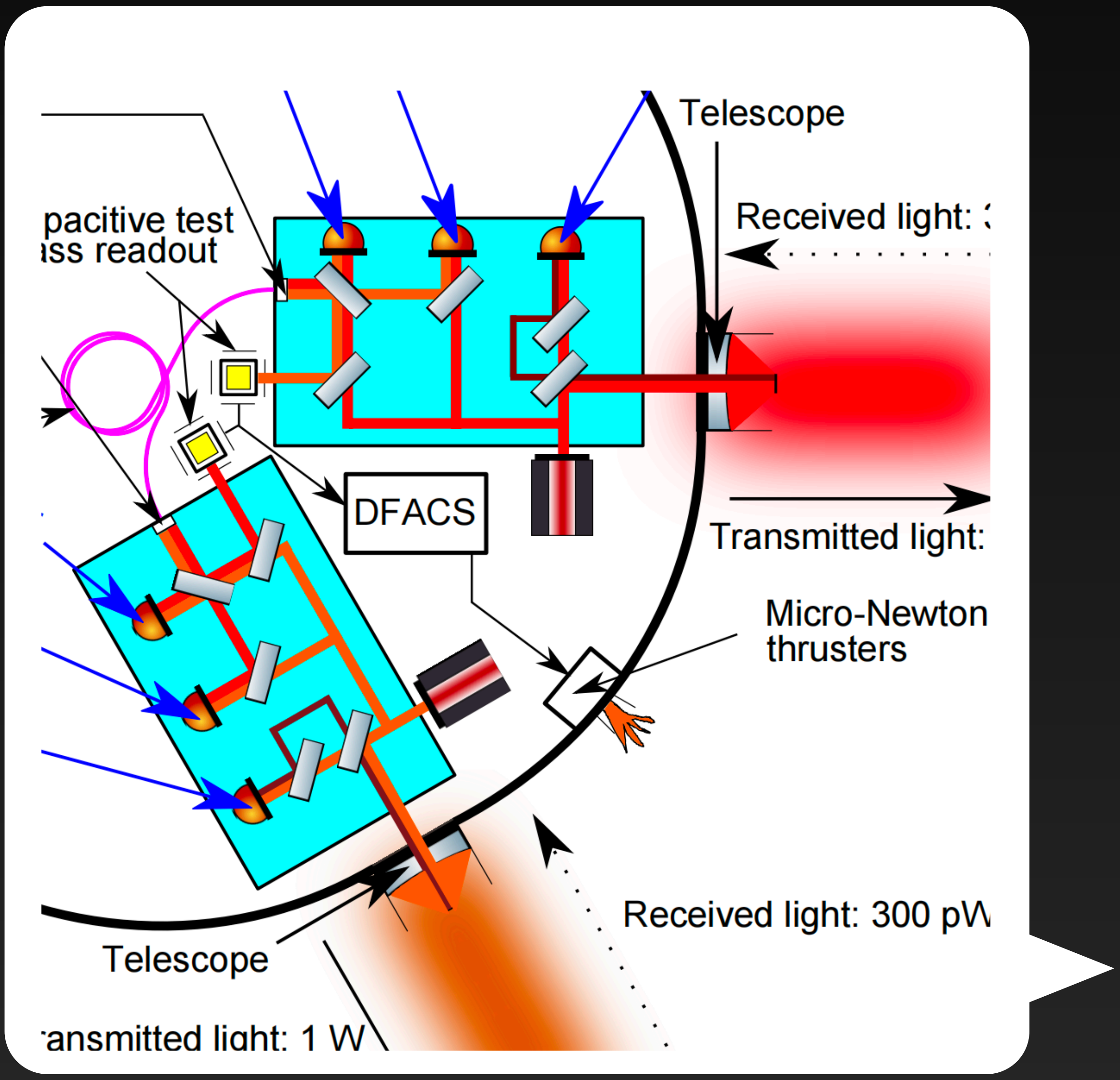
- GW scientists / data analyzers for space-based GW detections (LISA/Taiji/Tianqin)
- Researchers in the field of Continuous GW detection / time-frequency domain analysis
- Anyone who are curious about how is Taiji data different from LVK

1. Basic concepts for Taiji and space-based GW detection
2. How is Taiji's data different from LVK
3. The new and more "realistic" Taiji Data Challenge
4. Non-stationary noises in Taiji's long-duration observation
5. Working in the time-frequency domain:
    - Noise characterization
    - Galactic binary estimation
6. Taking-away messages

1. **Basic concepts for Taiji and space-based GW detection**
2. How is Taiji's data different from LVK
3. The new and more "realistic" Taiji Data Challenge
4. Non-stationary noises in Taiji's long-duration observation
5. Working in the time-frequency domain:
    - Noise characterization
    - Galactic binary estimation
6. Taking-away messages

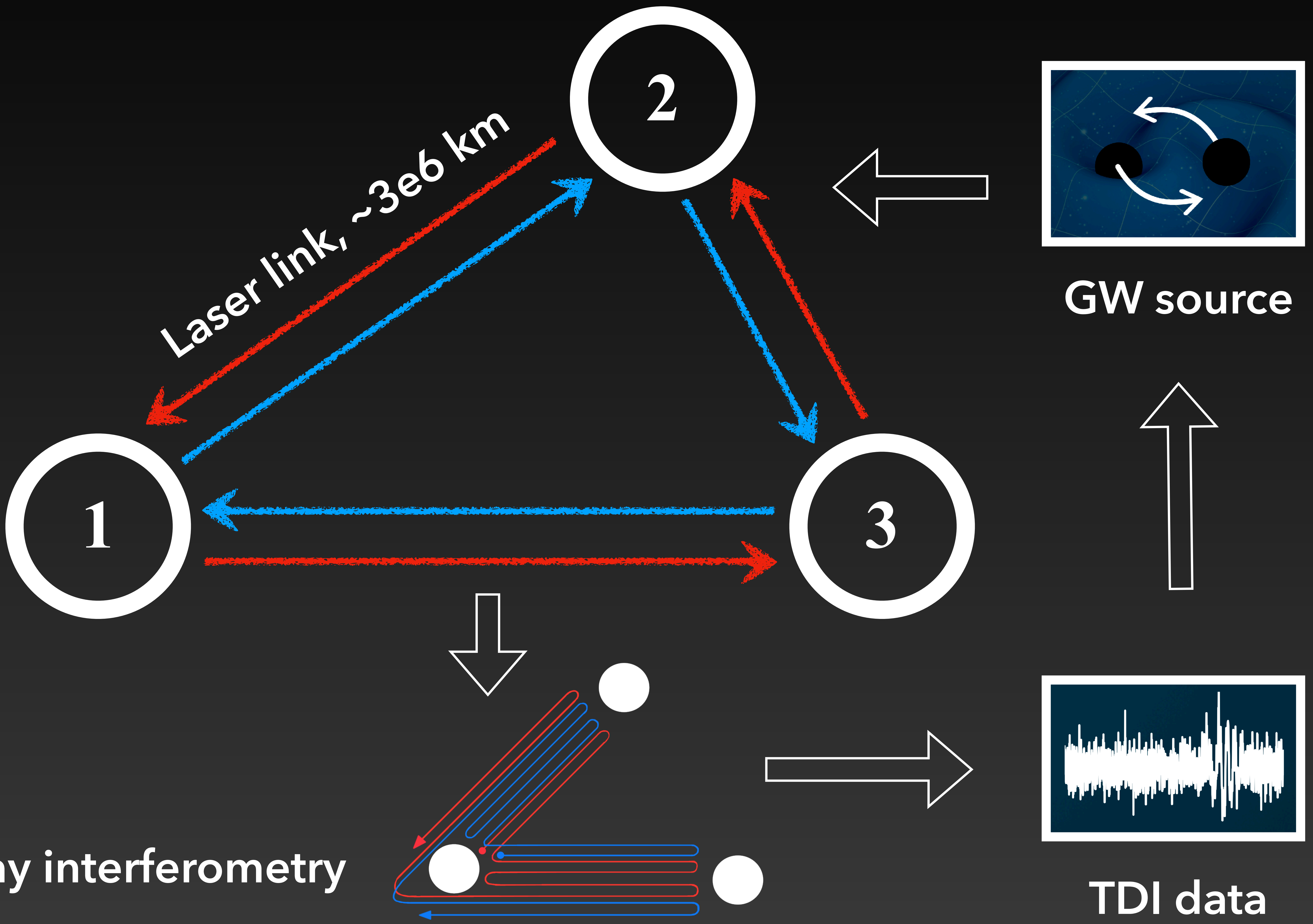
Basic Concepts: the Taiji Detector

[M. Otto, PhD thesis]



On each SC

- 2 identical test masses, serving as inertial references.
- laser interferometers to readout the changes of optical paths caused by GWs



Spacecraft (SC) (leads the Earth by approximately 20 degrees)

Laser link, ~3e6 km

GW source

Time-delay interferometry

TDI data

Basic Concepts: the Targets of Taiji

In > 5 year operation: burst, continuous, and stochastic GW signals in the mHz band

- Massive black hole binaries (MBHBs)

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

- Galactic compact binaries (GBs)

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

- Extreme mass-ratio inspirals (EMRIs)

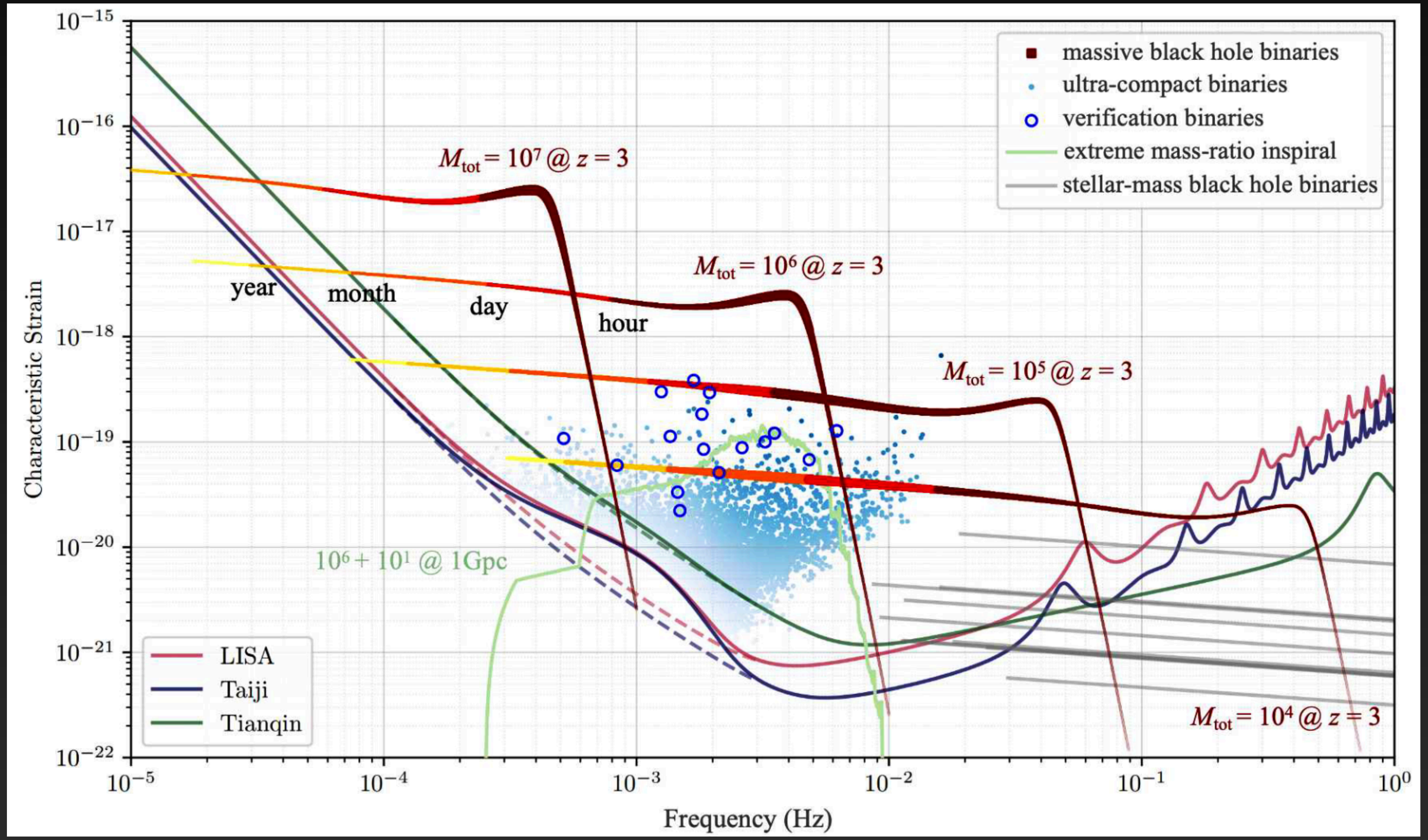
  $m_2/m_1 \in (10^{-3}, 10^{-6})$, $\mathcal{O}(1) \sim \mathcal{O}(10^3)$

- Stellar mass black hole binaries (SMBHBs)

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

- Stochastic GW background (SGWB)

- …



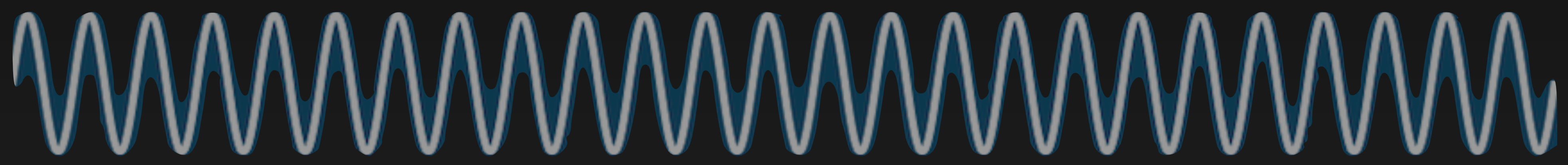
Target sources and sensitivities of LISA, Taiji, Tianqin

Basic Concepts: the Targets of Taiji

- Massive black hole binaries (MBHBs)

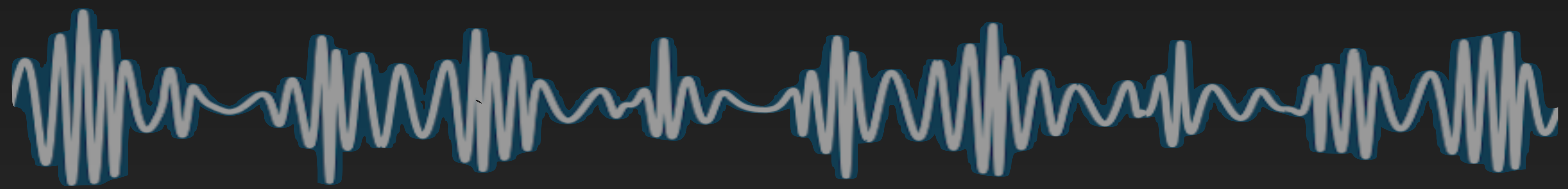
day-month

- Galactic binaries (GBs)



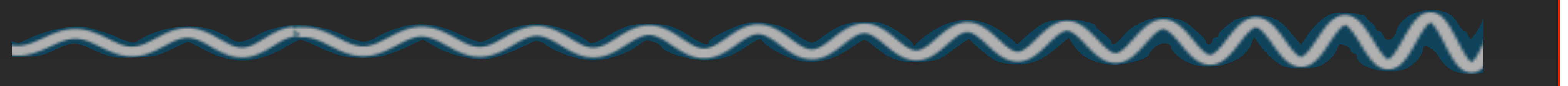
year

- Extreme mass-ratio inspirals (EMRIs)



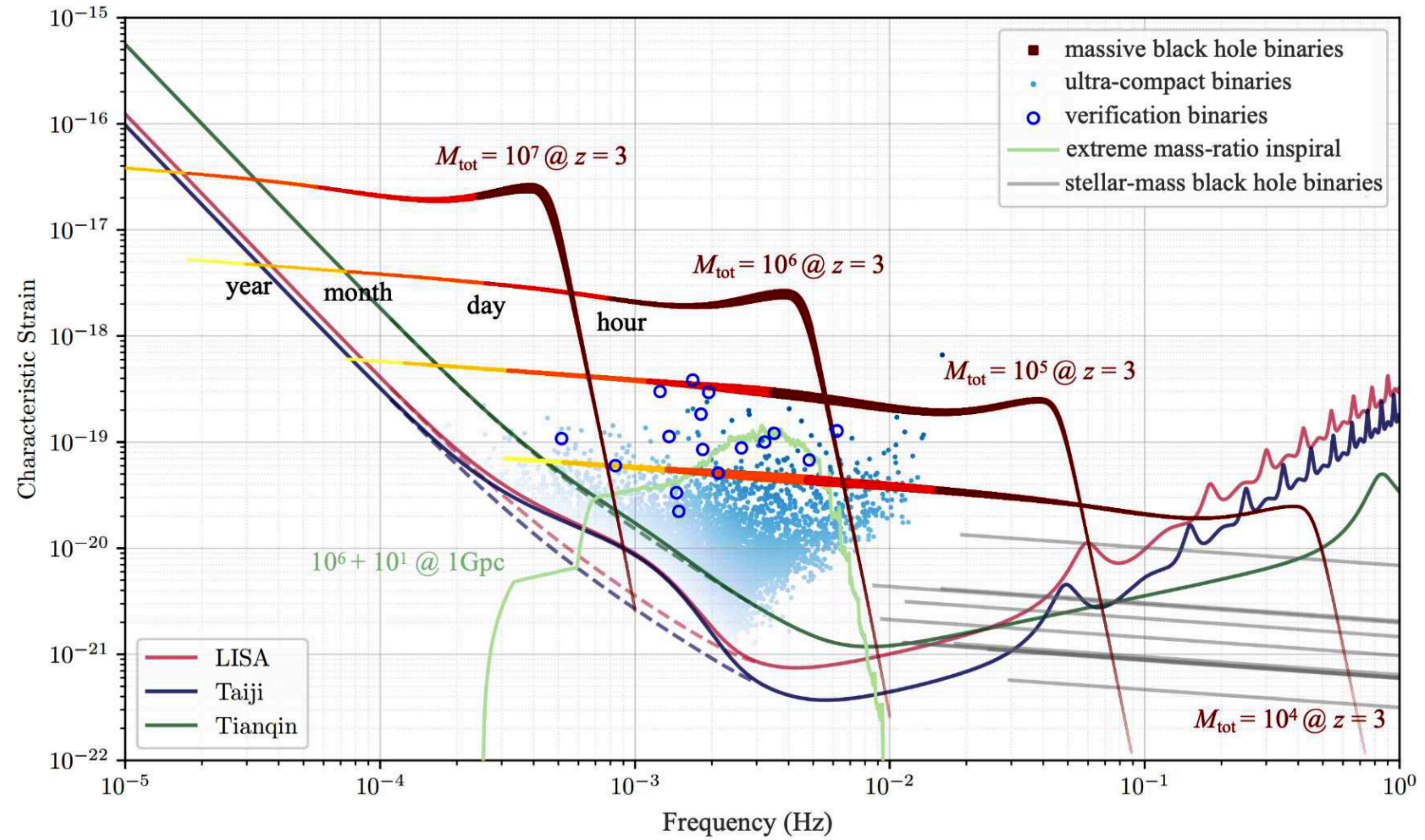
month-year

- Stellar mass black hole binaries (SMBHBs)



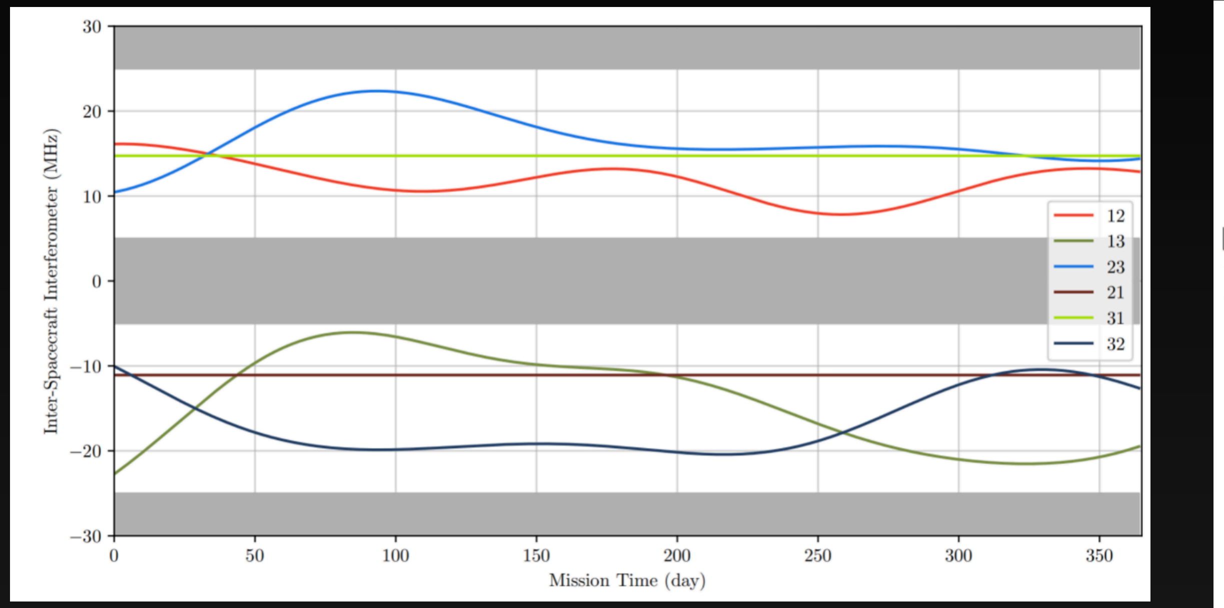
year

- Stochastic GW background (SGWB)
- …

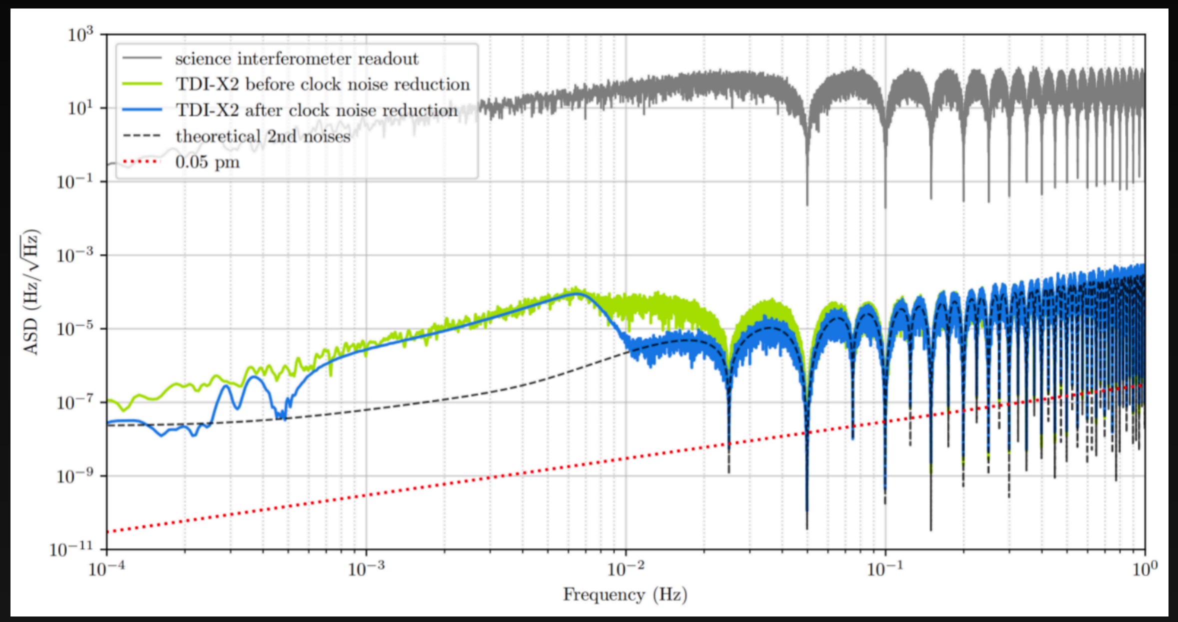


Target sources and sensitivities of LISA, Taiji, Tianqin

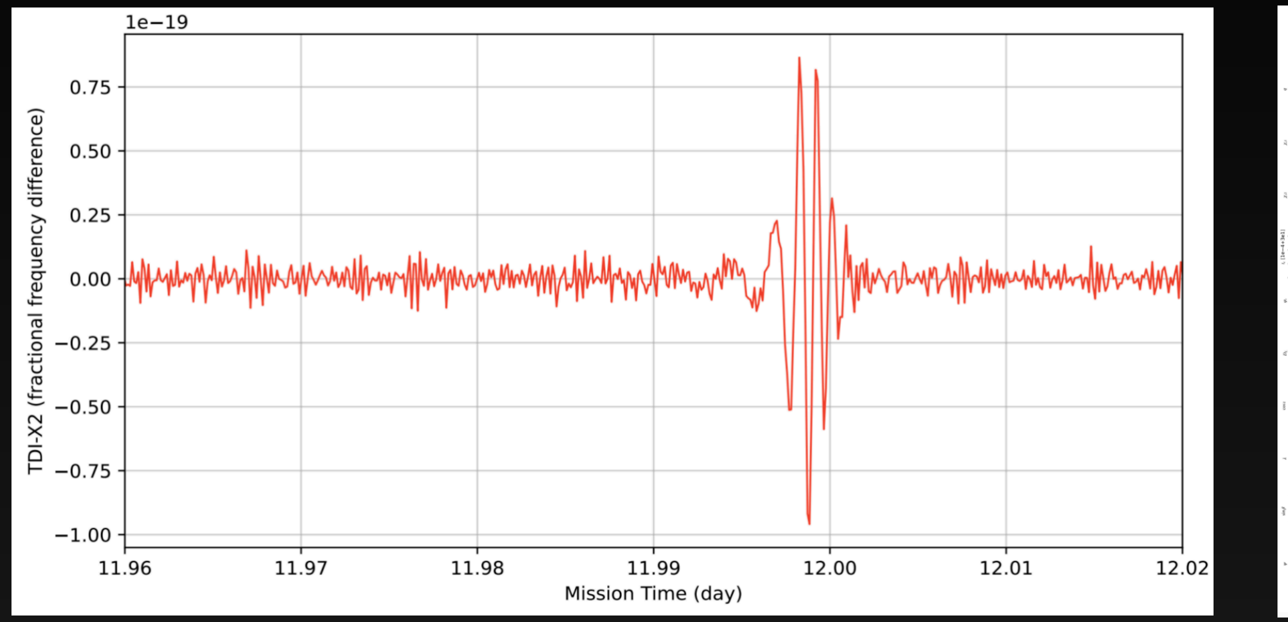
Basic Concepts: Taiji data flow



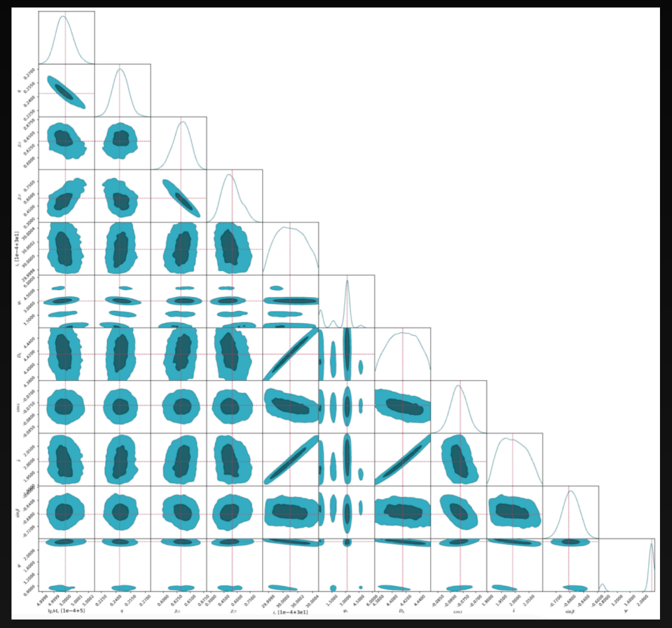
L0: Raw interferometric data



TDI processing



L1: Science data



L2,3: GW science

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

Provide data that are suitable for further analysis

- **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**

Include at least two critical pipelines

- **Low latency alert pipeline**

Rapidly detect and locate transient sources, early-warning for follow-up GW and electro-magnetic observations

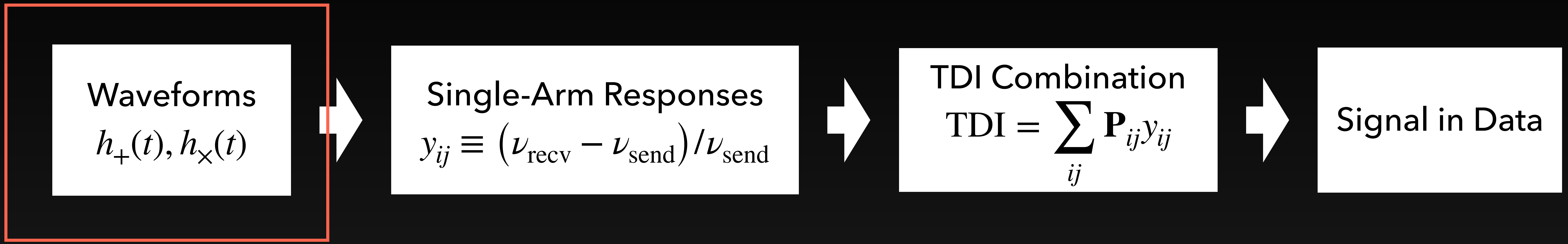
- **Global fit analysis pipeline**

Taiji data will be signal-dominated, numerous signals overlap in both the time domain and frequency domain. Indispensable to jointly estimate the parameters of thousands of sources

- **Further scientific interpretations**

1. Basic concepts for Taiji and space-based GW detection
2. **How is Taiji’s data different from LVK**
3. The new and more “realistic” Taiji Data Challenge
4. Non-stationary noises in Taiji’s long-duration observation
5. Working in the time-frequency domain:
    - Noise characterization
    - Galactic binary estimation
6. Taking-away messages

How is Taiji data different from LVK?

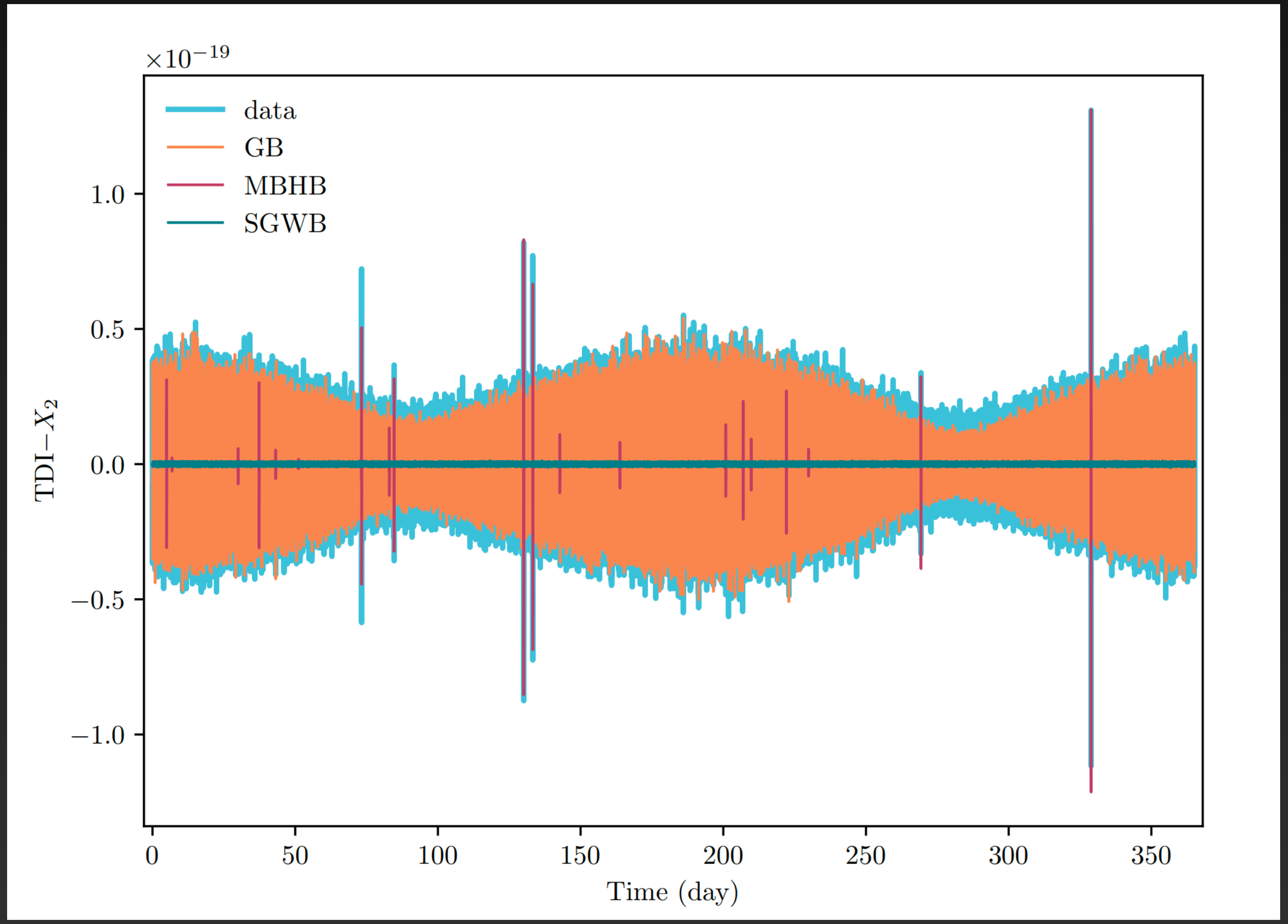


Waveforms
$h_+(t), h_\times(t)$

Single-Arm Responses
$y_{ij} \equiv \left(\nu_{\text{recv}} - \nu_{\text{send}}\right)/\nu_{\text{send}}$

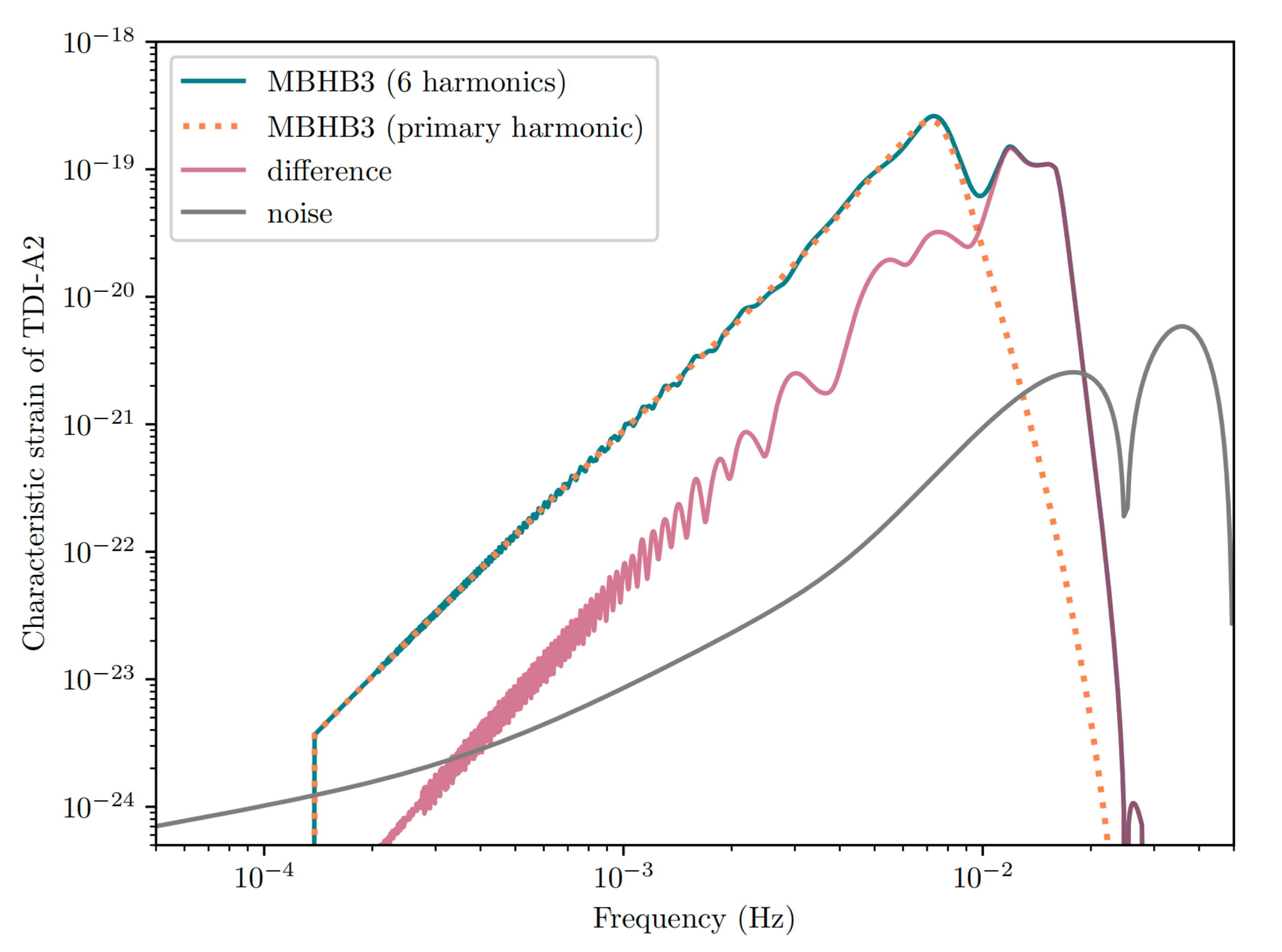
TDI Combination
$\text{TDI} = \sum_{ij} \mathbf{P}_{ij} y_{ij}$

Signal in Data



Time scales: days - months - years

majority of detectable signals would be continuous GWs



High accuracy requirement in the high-SNR scenario

Require more accurate waveform templates to make unbiased estimate, subdominant features can not be neglected: higher harmonics, eccentricity, … current template models may become insufficient

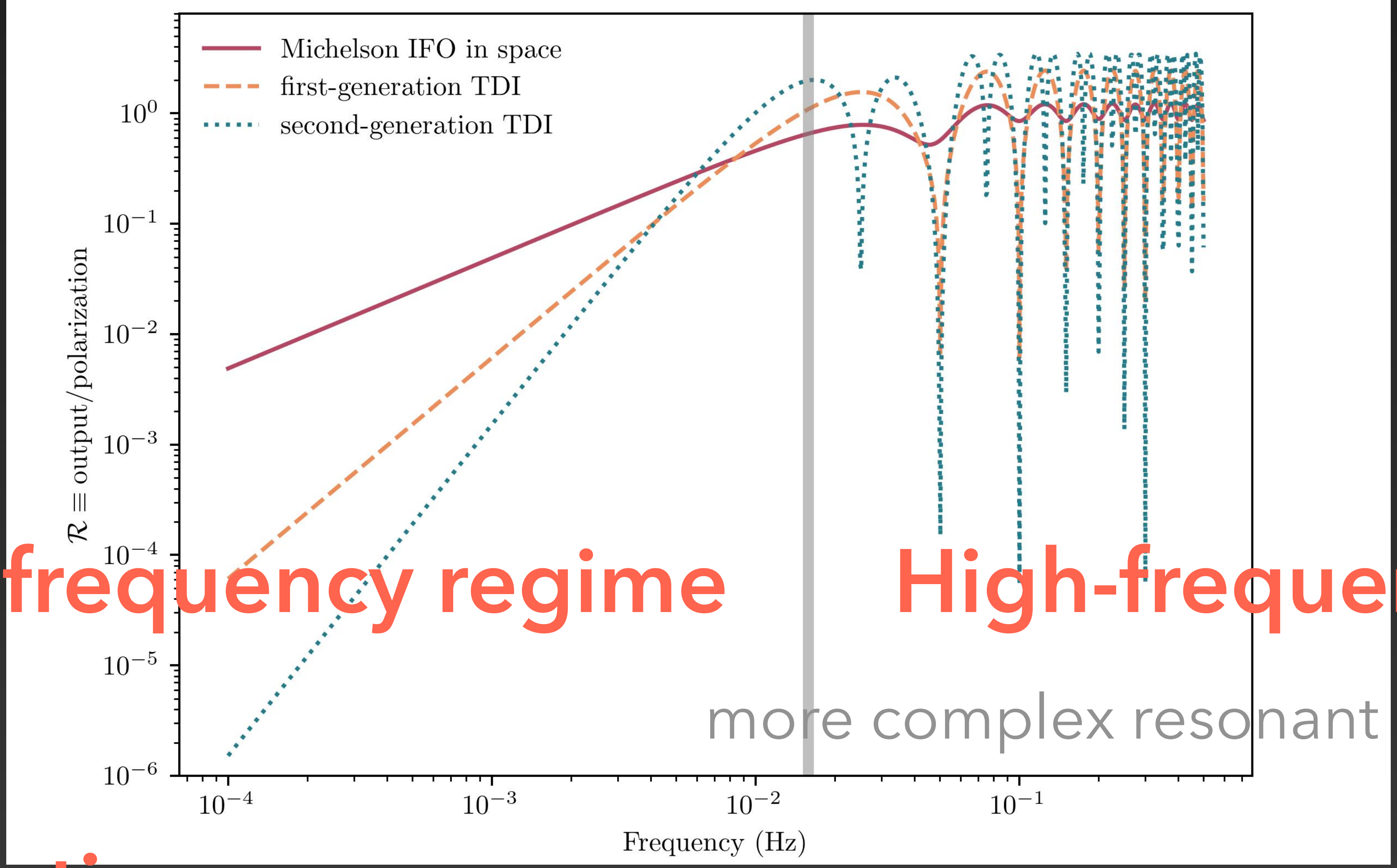
How is Taiji data different from LVK?

Waveforms $h_+(t), h_\times(t)$ → Single-Arm Responses $y_{ij} \equiv \left(\nu_{\mathrm{recv}} - \nu_{\mathrm{send}}\right)/\nu_{\mathrm{send}}$ → TDI Combination $\mathrm{TDI} = \sum_{ij} \mathbf{P}_{ij} y_{ij}$ → Signal in Data

where orbit information enters

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

1. Long-wavelength approximation does not always apply
2. Effects of orbital dynamics: Doppler, yearly amplitude modulation, variation of arms …



Low-frequency regime

High-frequency regime

more complex resonant behavior appears

Not anymore time-independent antenna pattern function

How is Taiji data different from LVK?

Waveforms
$h_+(t), h_\times(t)$

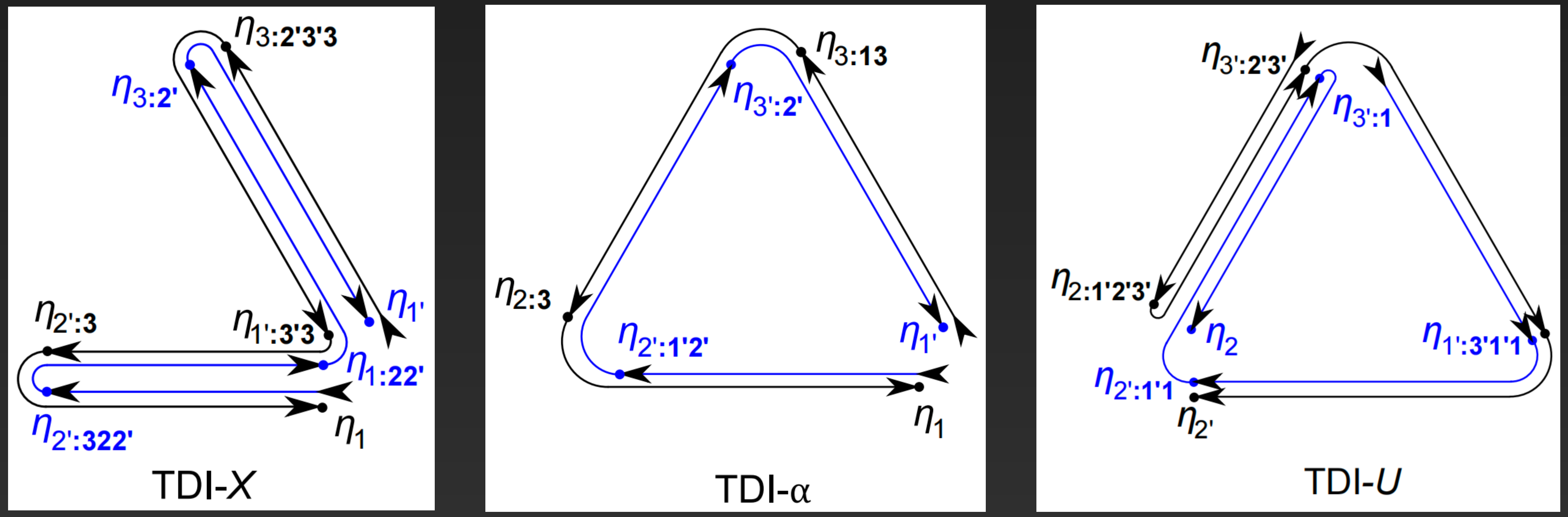
Single-Arm Responses
$y_{ij} \equiv \left(\nu_{\text{recv}} - \nu_{\text{send}}\right)/\nu_{\text{send}}$

TDI Combination
$\text{TDI} = \sum_{ij} \mathbf{P}_{ij} y_{ij}$

Signal in Data

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

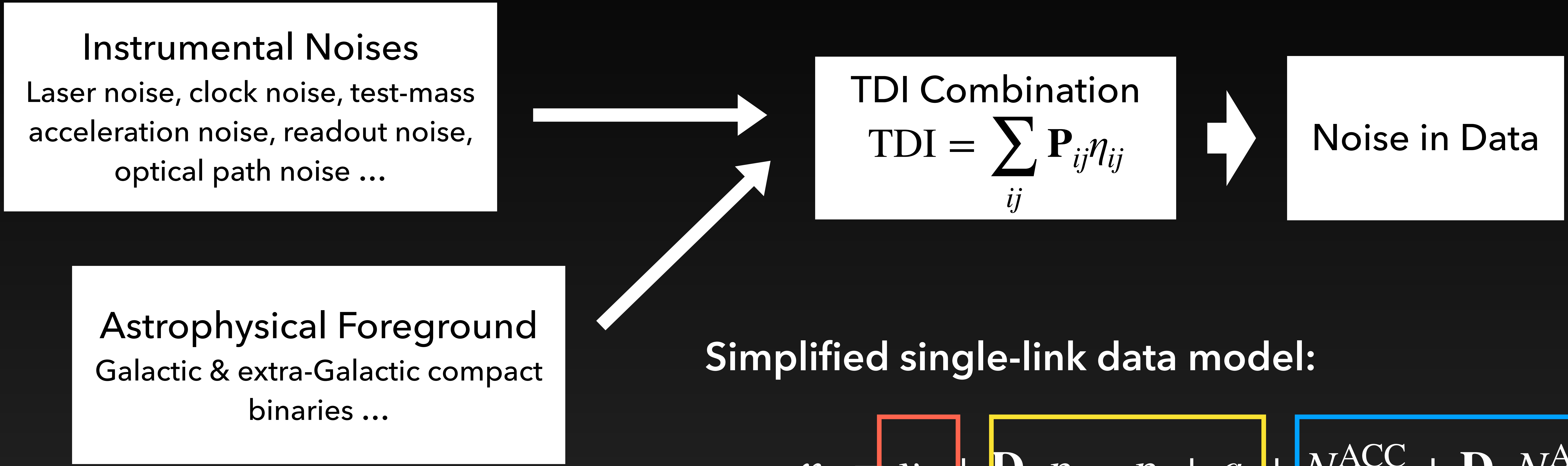
where orbit information enters



Different TDI configurations, 3 are shown for example
[M. Otto, PhD thesis]

1. Various TDI configurations
2. The combinations are also affected by the length of arms (delays)

How is Taiji data different from LVK?



Instrumental Noises
Laser noise, clock noise, test-mass acceleration noise, readout noise, optical path noise …

Astrophysical Foreground
Galactic & extra-Galactic compact binaries …

TDI Combination

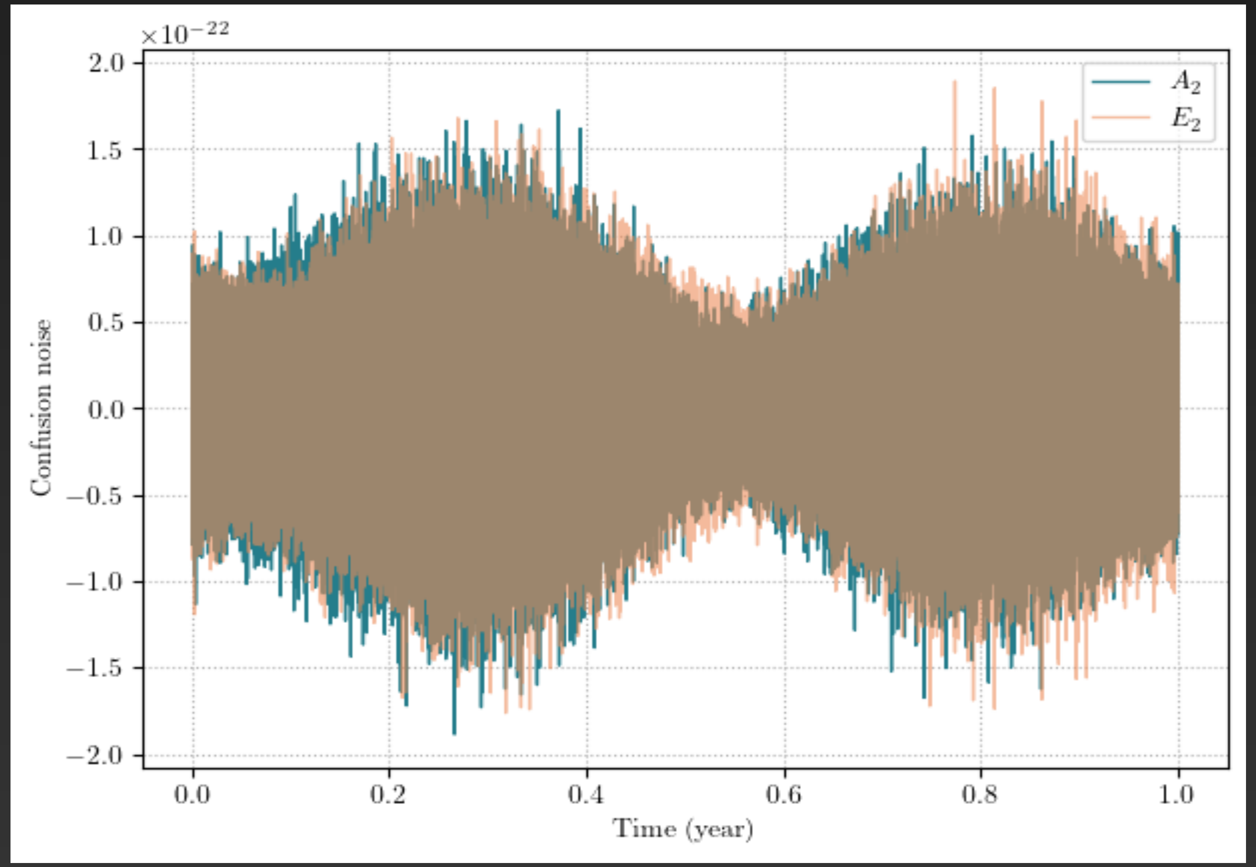
$$\mathrm{TDI} = \sum_{ij} \mathbf{P}_{ij}\eta_{ij}$$

Noise in Data

Simplified single-link data model:

$$\eta_{ij} = \boxed{y_{ij}} + \boxed{\mathbf{D}_{ij}p_j - p_i + q_i} + \boxed{N^{\mathrm{ACC}}_{ij} + \mathbf{D}_{ij}N^{\mathrm{ACC}}_{ji} + N^{\mathrm{OMS}}_{ij}}$$

GW signal suppressed by TDI left in data



Galactic foreground

1. Both of them remain incompletely understood, and we may not be able to monitor them with redundant auxiliary sensors
2. Non-stationary in the observation of year-long signals

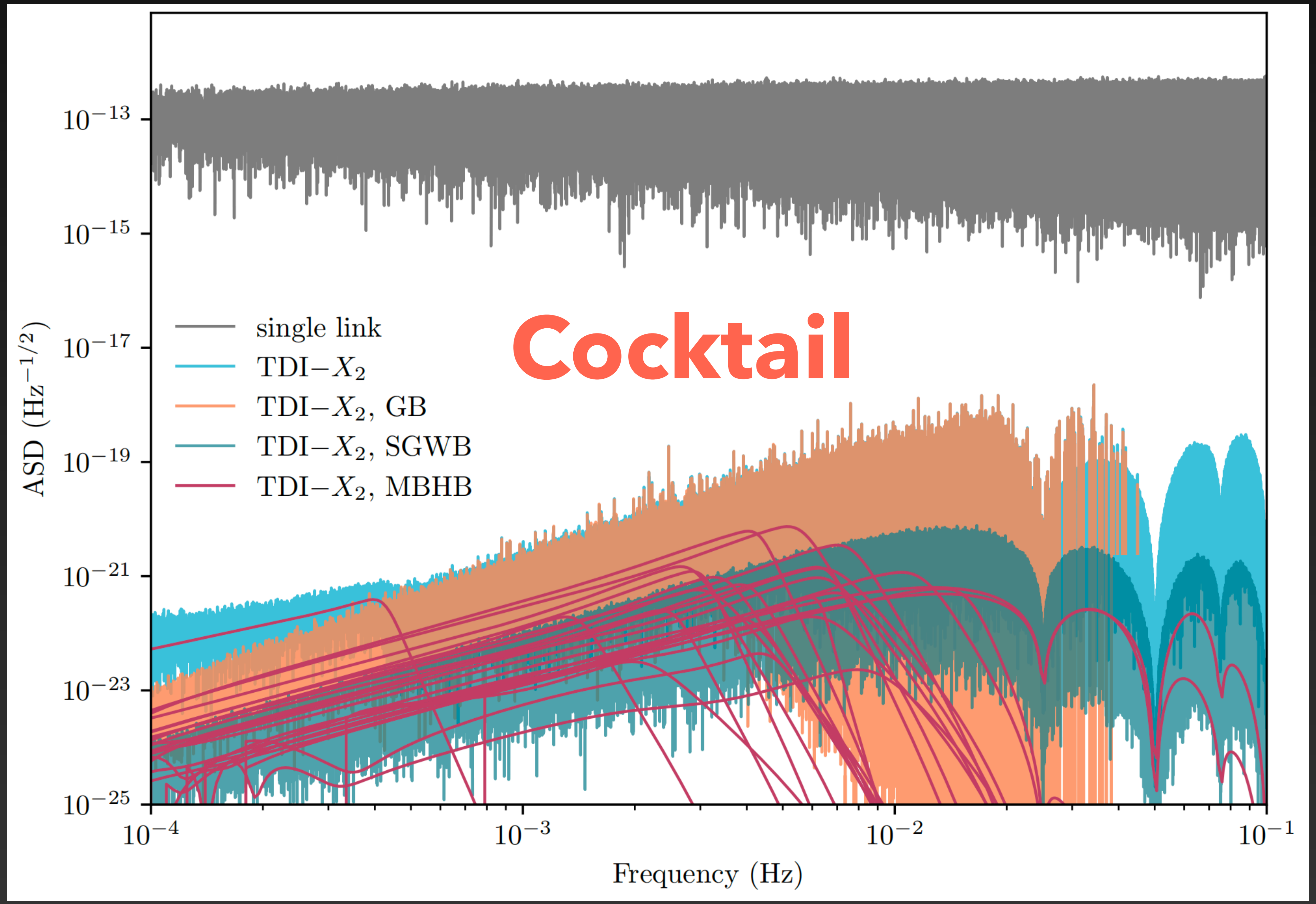
(a possible solution later)

How is Taiji data different from LVK?

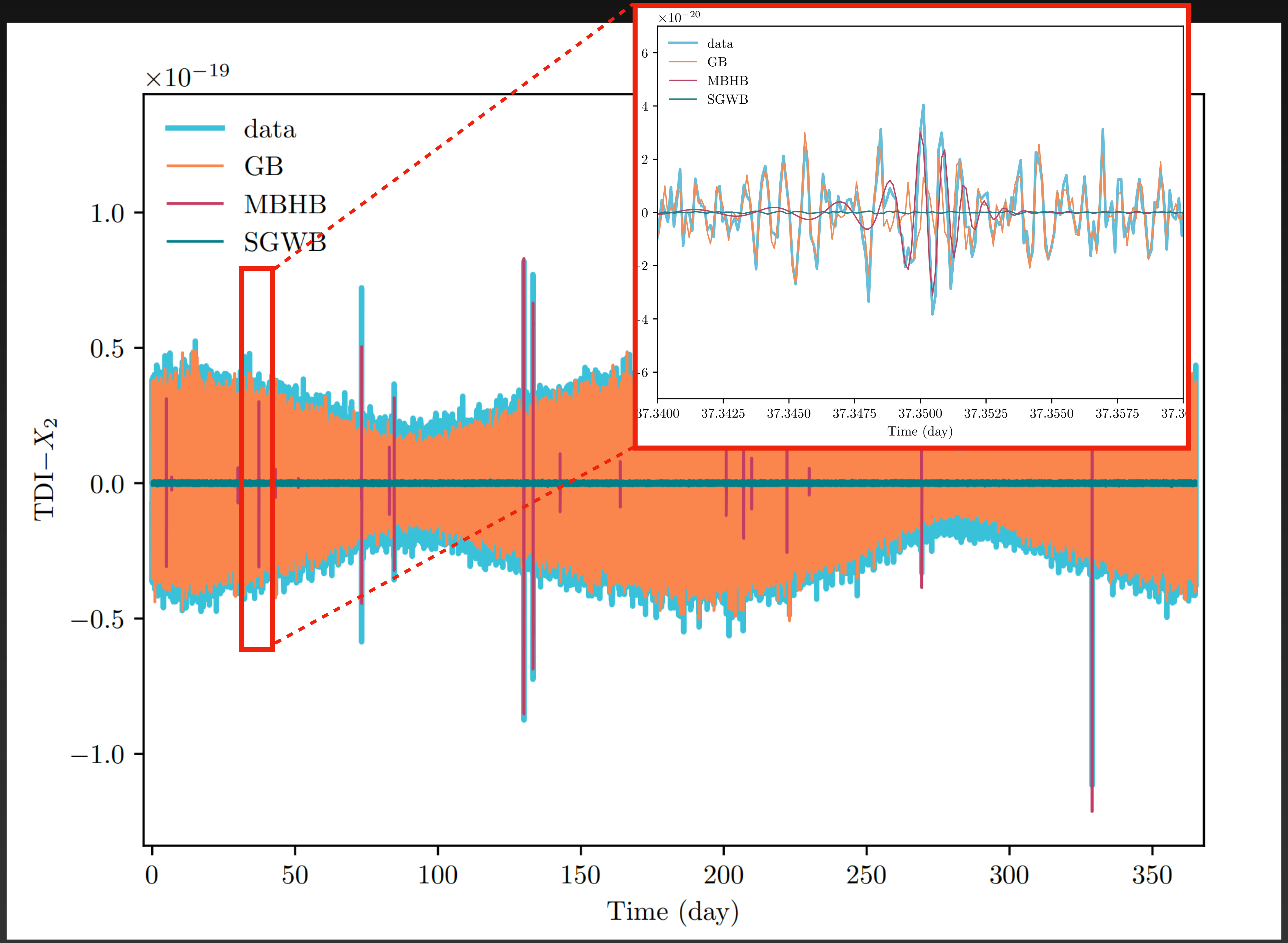
Core challenge of Taiji data analysis:

cocktail ——— the mixture of noises and numerous overlapping signals

Signals overlap in both time domain and frequency domain, we need a global analysis pipeline

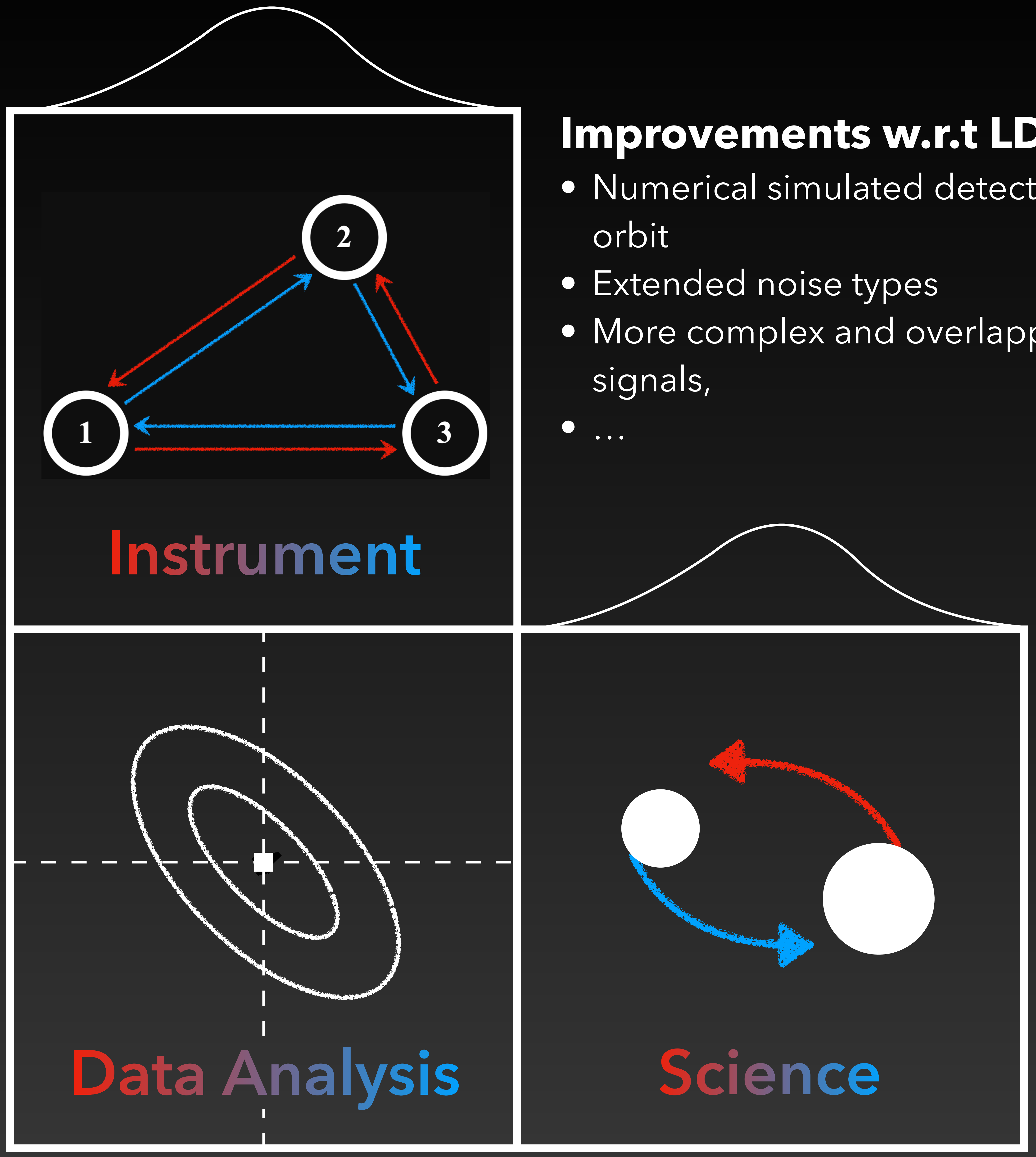


Data in frequency domain (ASD)



Data in time domain (downsampled to 0.1 Hz)

1. Basic concepts for Taiji and space-based GW detection
2. How is Taiji's data different from LVK
3. **The new and more "realistic" Taiji Data Challenge**
4. Non-stationary noises in Taiji's long-duration observation
5. Working in the time-frequency domain:
    - Noise characterization
    - Galactic binary estimation
6. Taking-away messages



Improvements w.r.t LDCs:

- Numerical simulated detector orbit
- Extended noise types
- More complex and overlapping signals,
- …

The TDC working group

Taiji Data Challenge

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

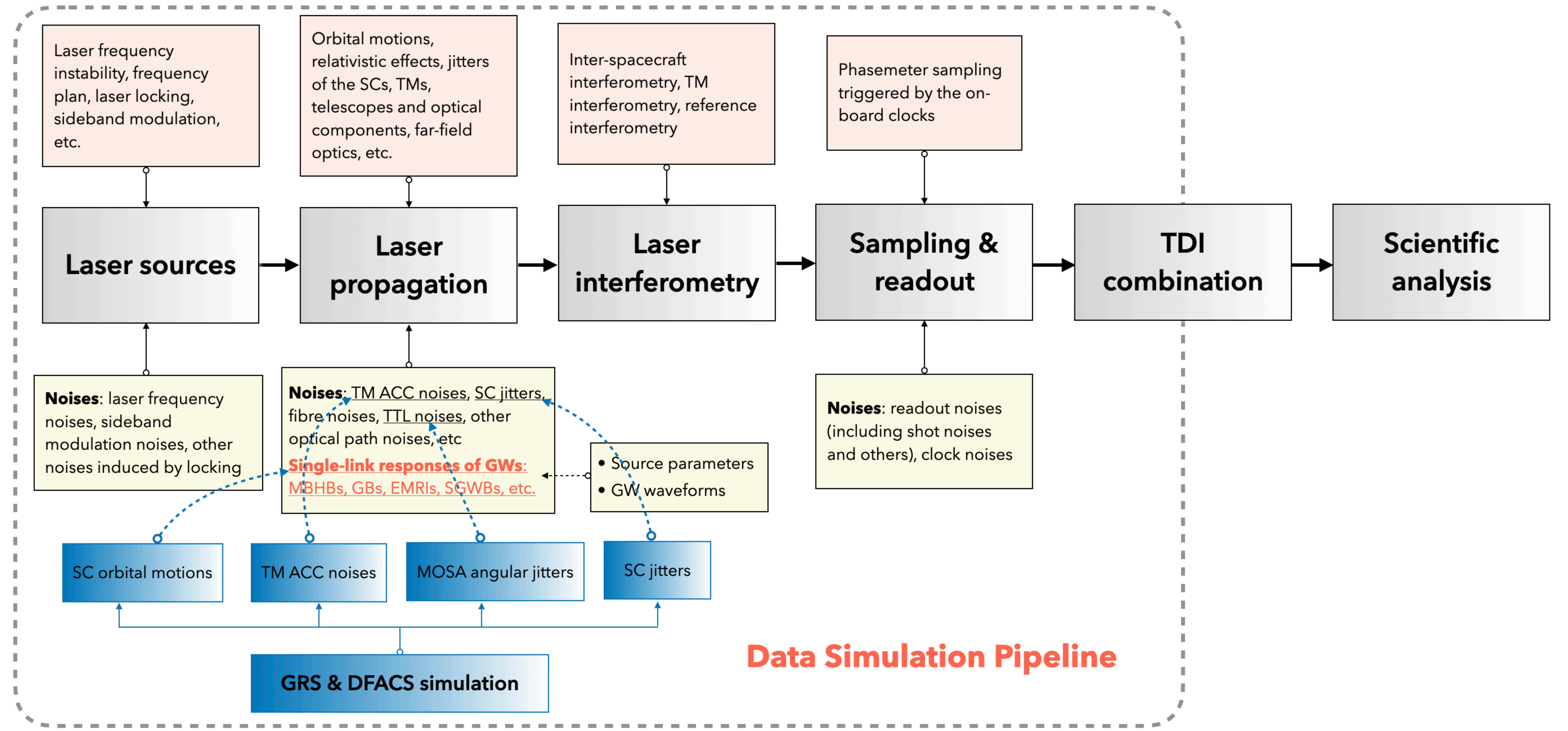
Available after May, 23, 2025

Website: http://gr.imech.ac.cn/overview/

GitHub: https://github.com/TriangleDataCenter

- Download links of Data
- Manual, Presentation: components, creation …
- Tools and Tutorials help to access and use data

Data Simulation Pipeline



Laser frequency instability, frequency plan, laser locking, sideband modulation, etc.

Orbital motions, relativistic effects, jitters of the SCs, TMs, telescopes and optical components, far-field optics, etc.

Inter-spacecraft interferometry, TM interferometry, reference interferometry

Phasemeter sampling triggered by the on-board clocks

Laser sources

Laser propagation

Laser interferometry

Sampling & readout

TDI combination

Scientific analysis

Noises: laser frequency noises, sideband modulation noises, other noises induced by locking

Noises: TM ACC noises, SC jitters, fibre noises, TTL noises, other optical path noises, etc

Single-link responses of GWs: MBHBs, GBs, EMRIs, SGWBs, etc.

- Source parameters
- GW waveforms

Noises: readout noises (including shot noises and others), clock noises

SC orbital motions

TM ACC noises

MOSA angular jitters

SC jitters

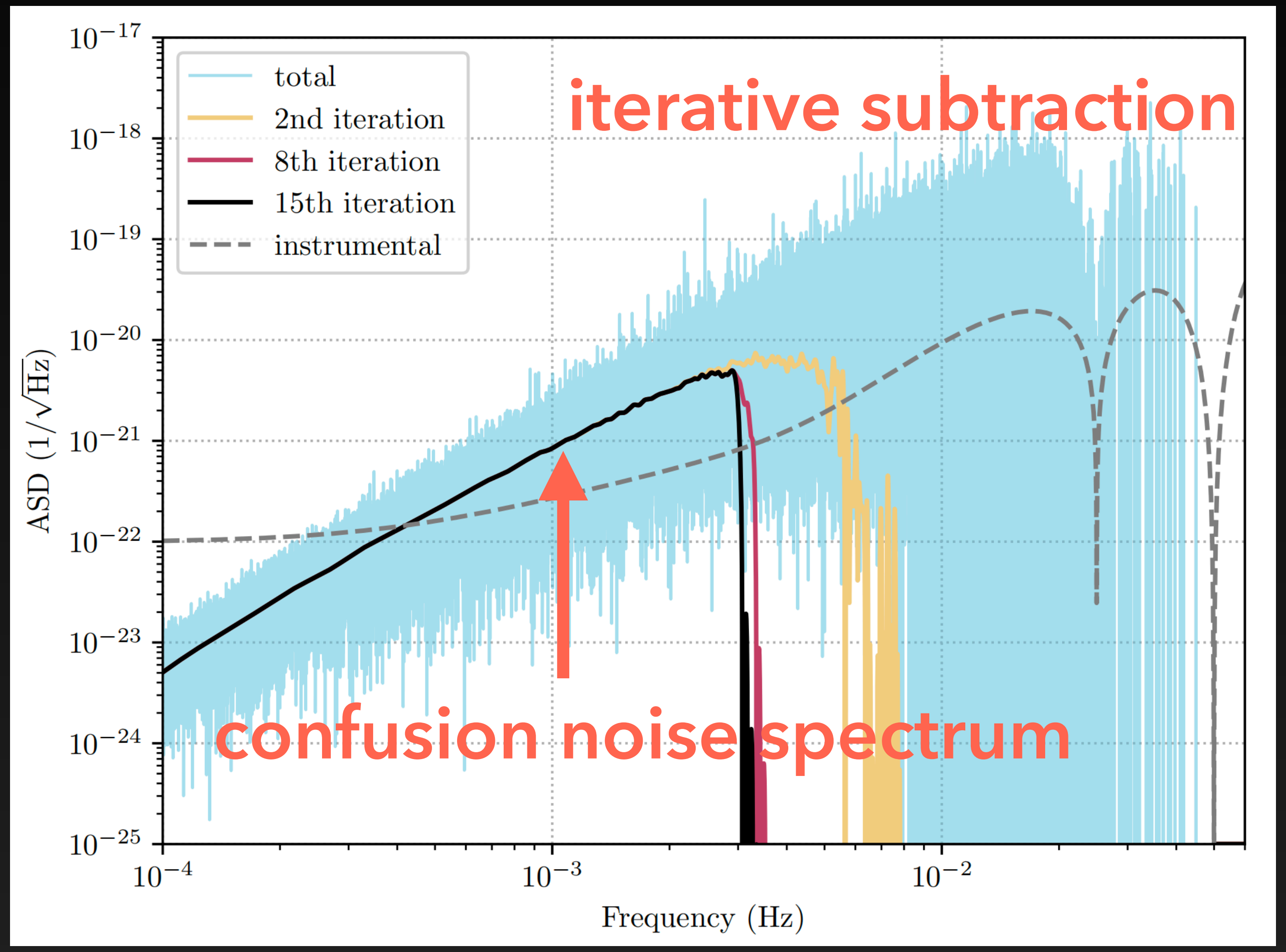
GRS & DFACS simulation

Data Simulation Pipeline

Models Behind the TDC Data

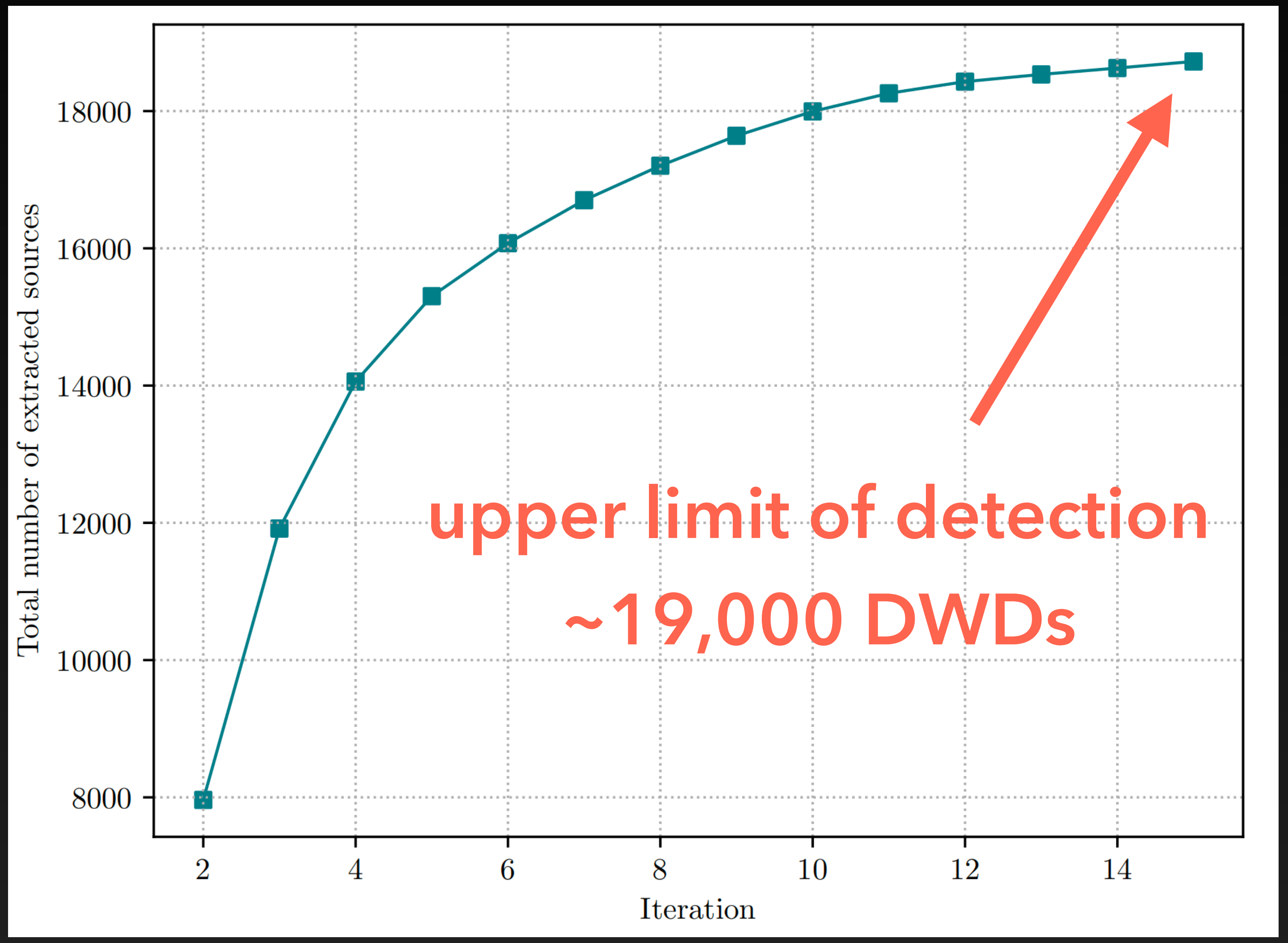
The TDC Populations

- **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

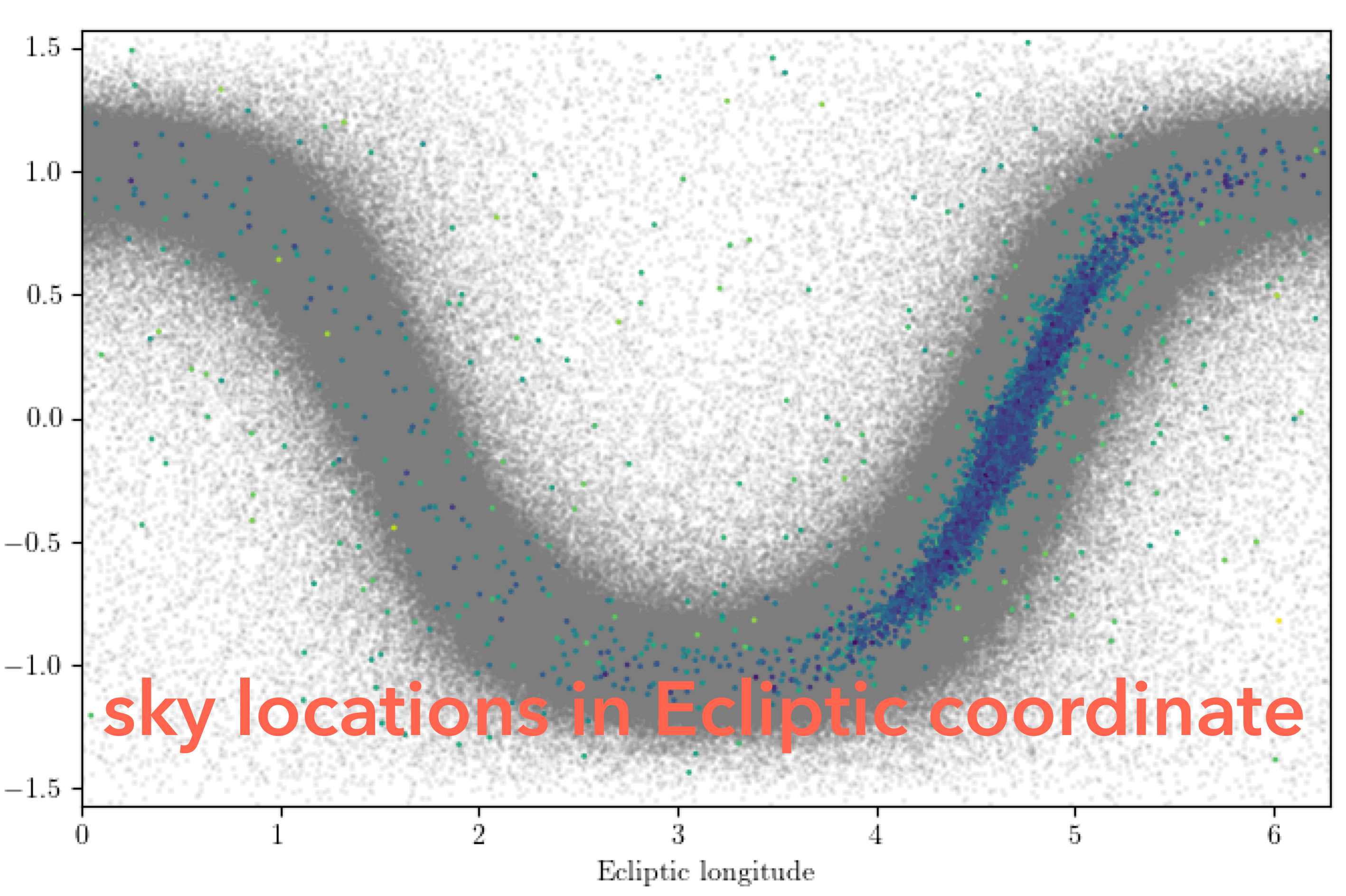


iterative subtraction

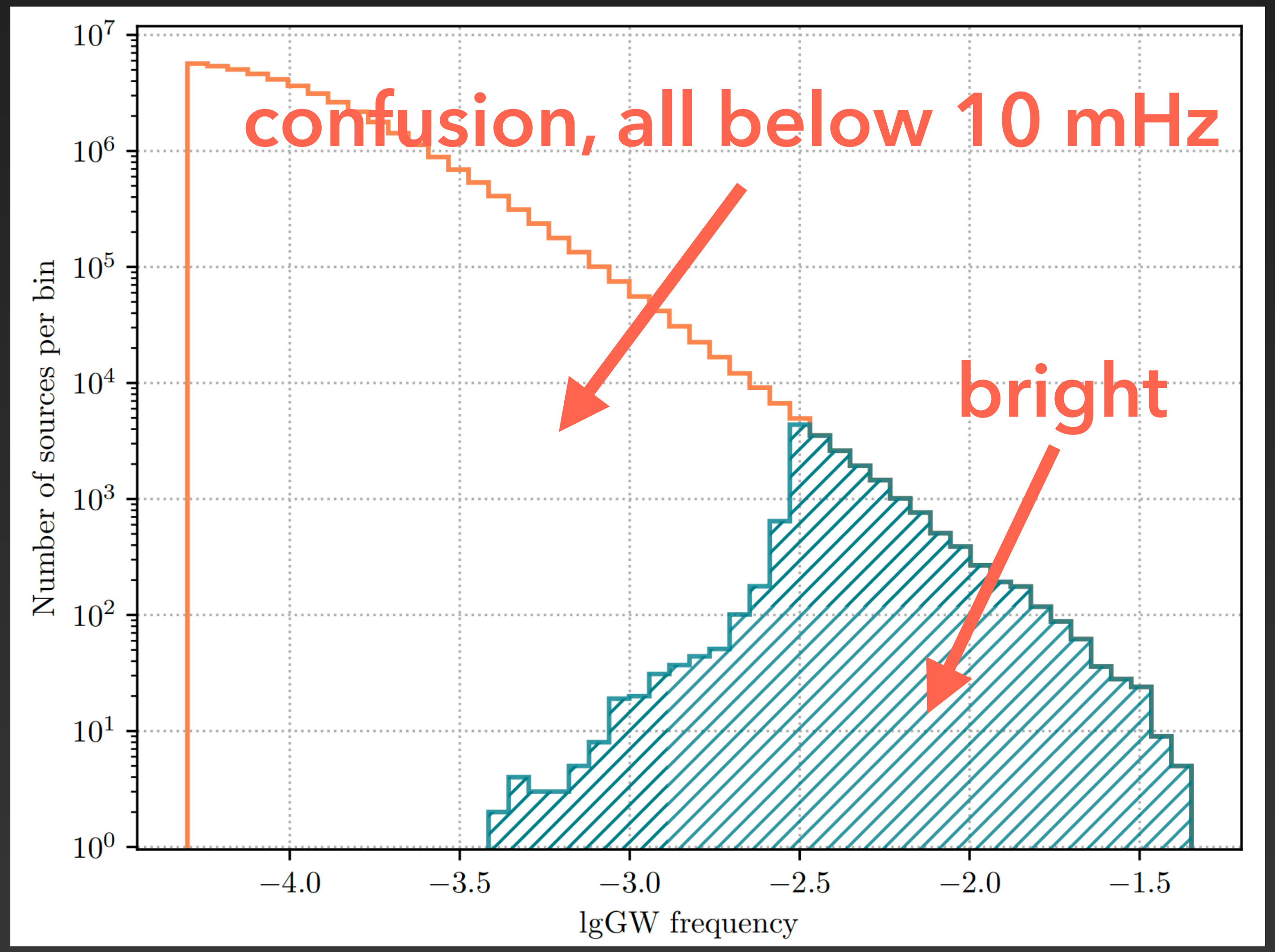
confusion noise spectrum



upper limit of detection ~19,000 DWDs



sky locations in Ecliptic coordinate



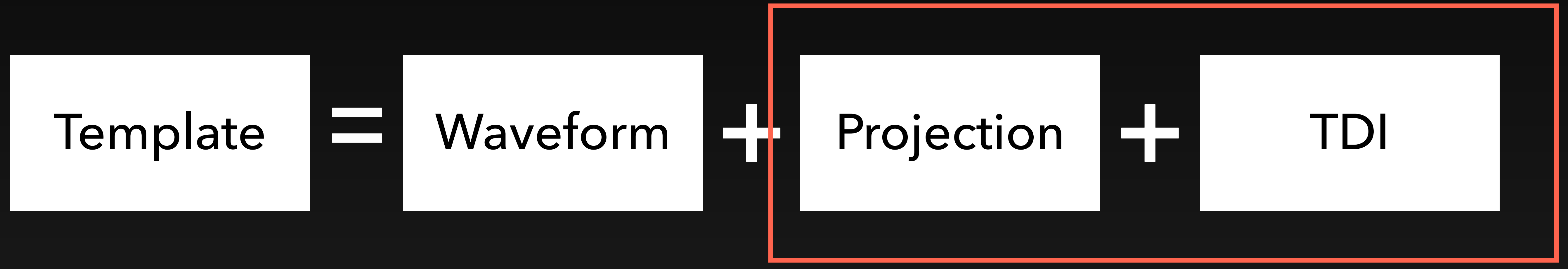
confusion, all below 10 mHz

bright

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

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

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

where orbit information enters

Models Behind the TDC Data

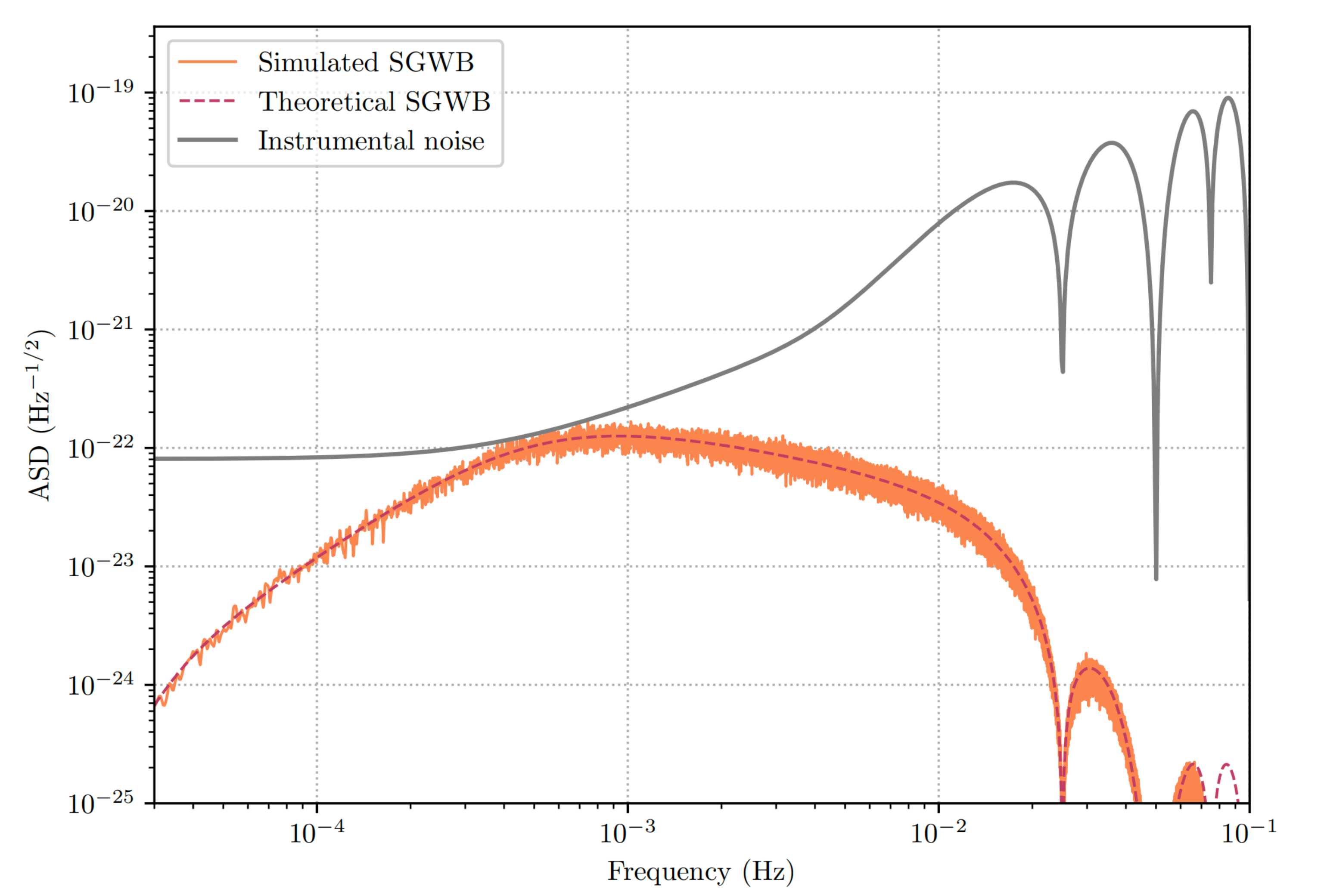
The TDC Waveforms

- **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

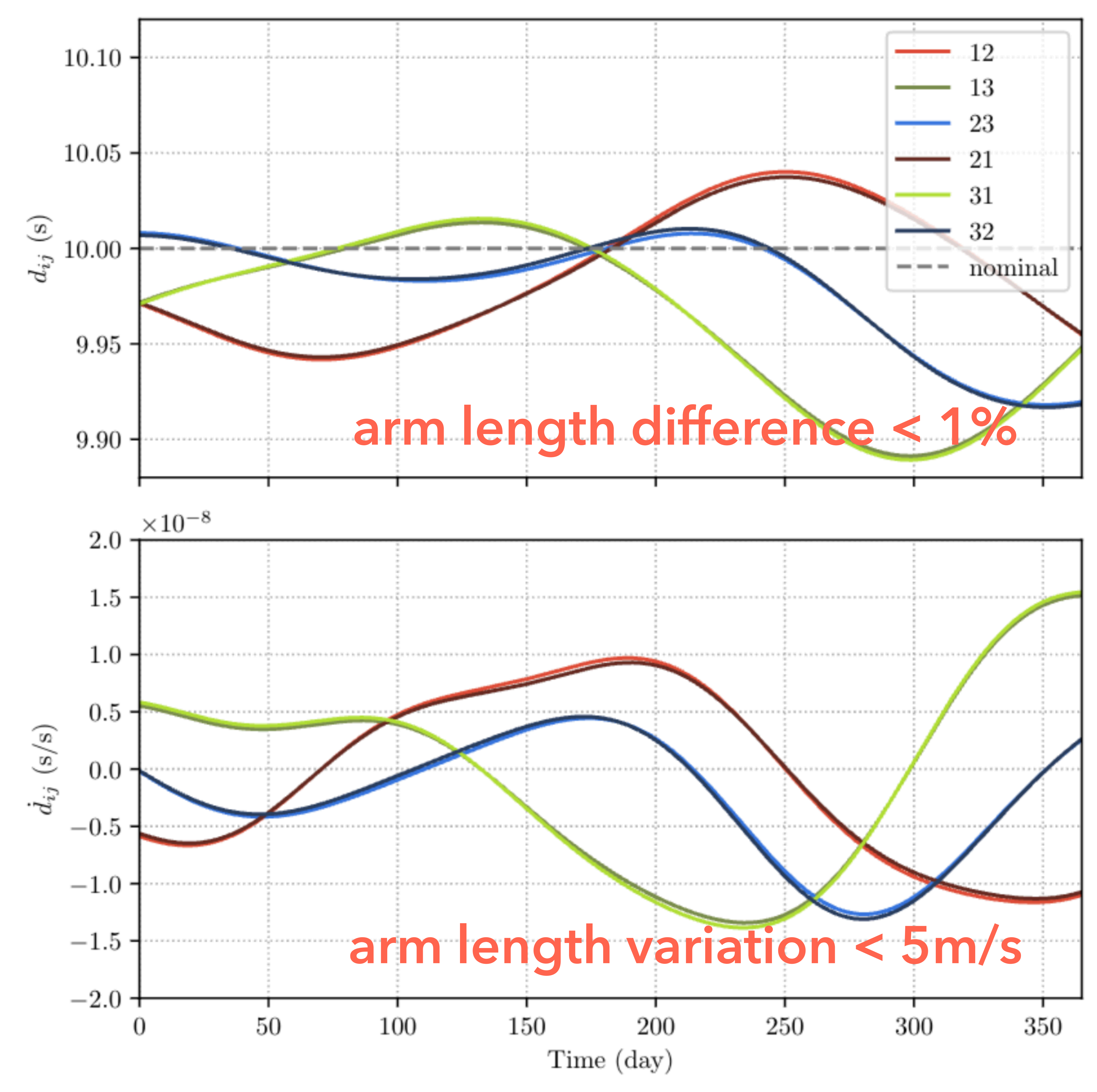
Slow, but can account for time-varying response & anisotropy SGWBs



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



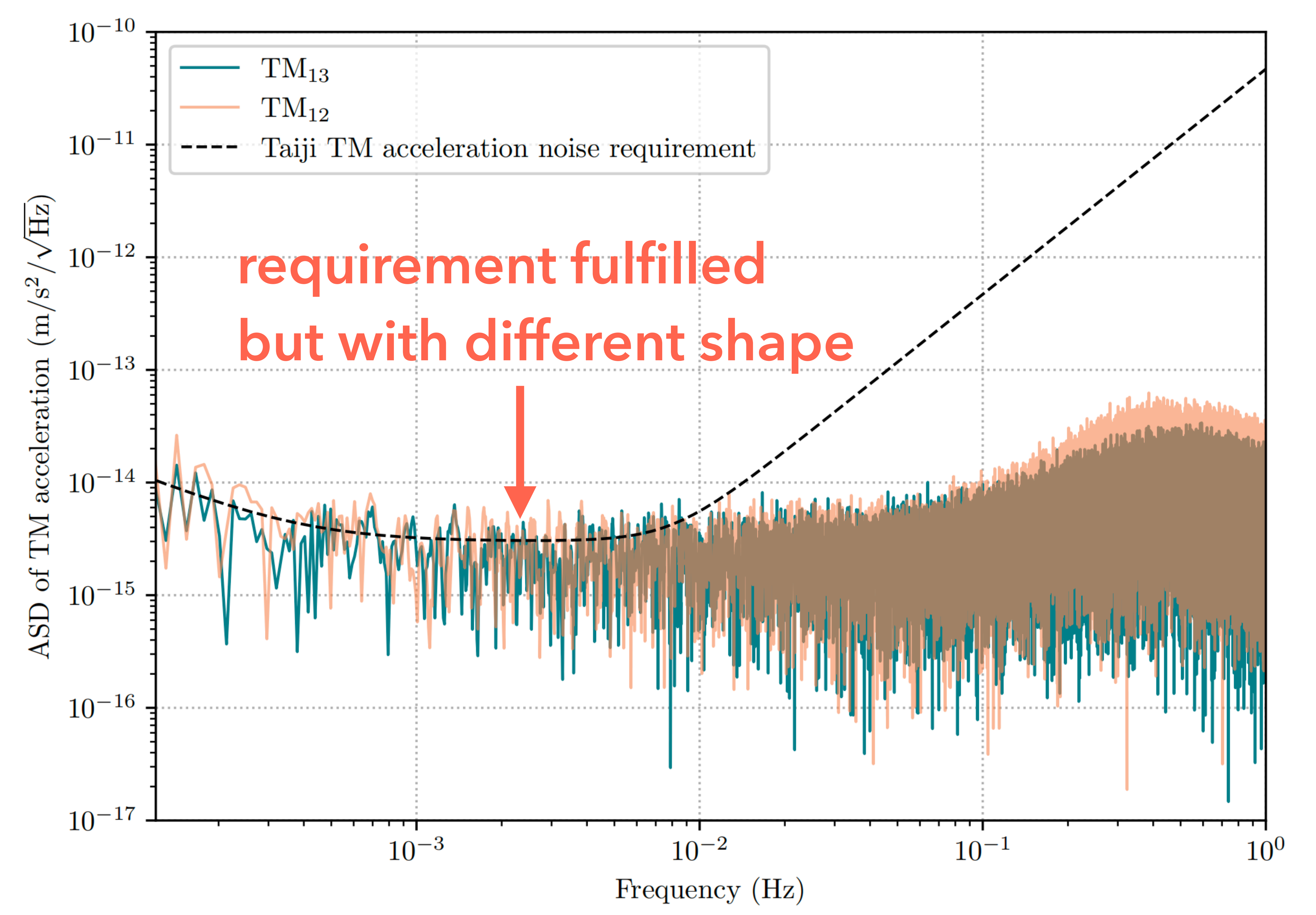
arm length difference < 1%

arm length variation < 5m/s

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



requirement fulfilled
but with different shape

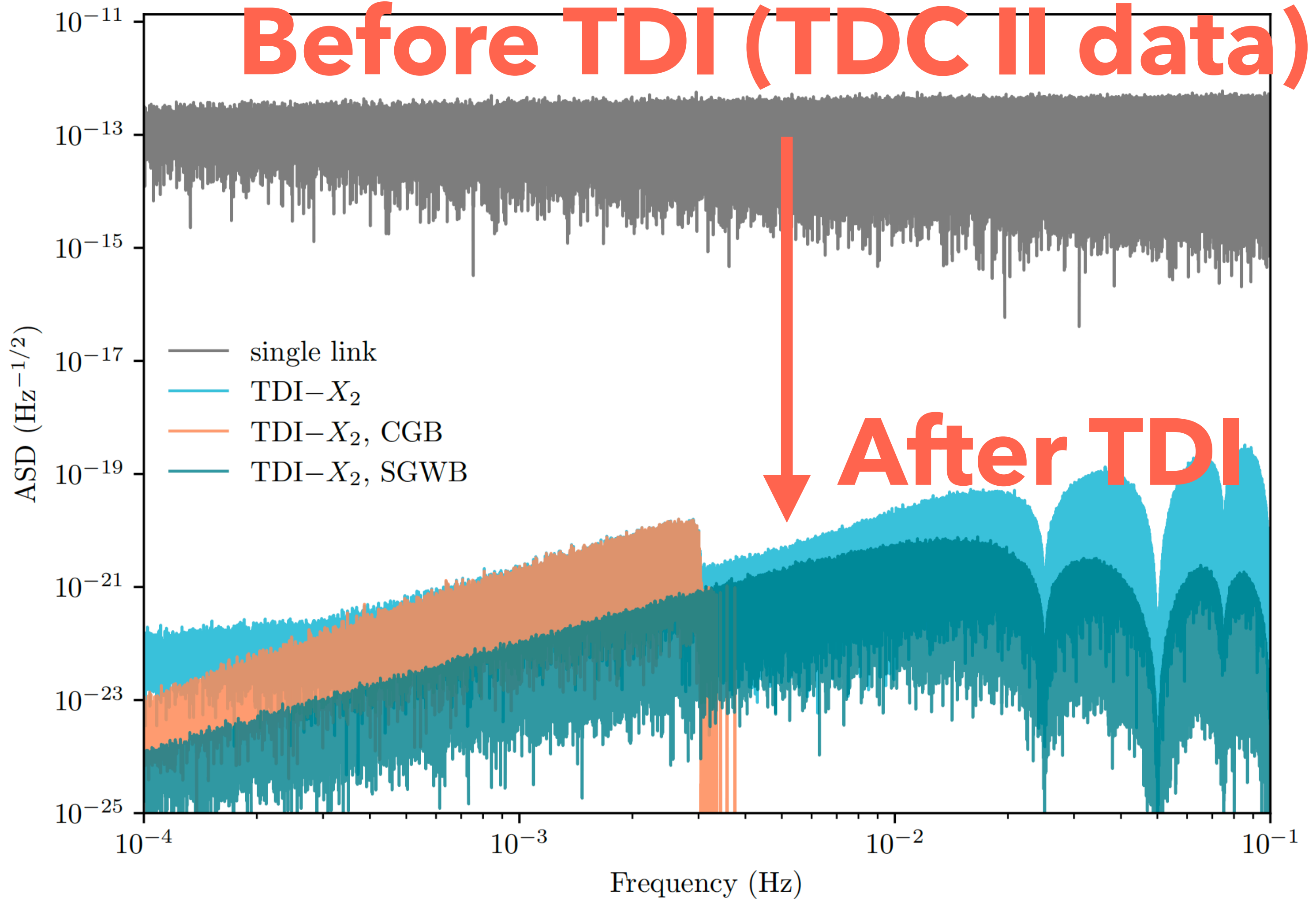
Models Behind the TDC Data

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

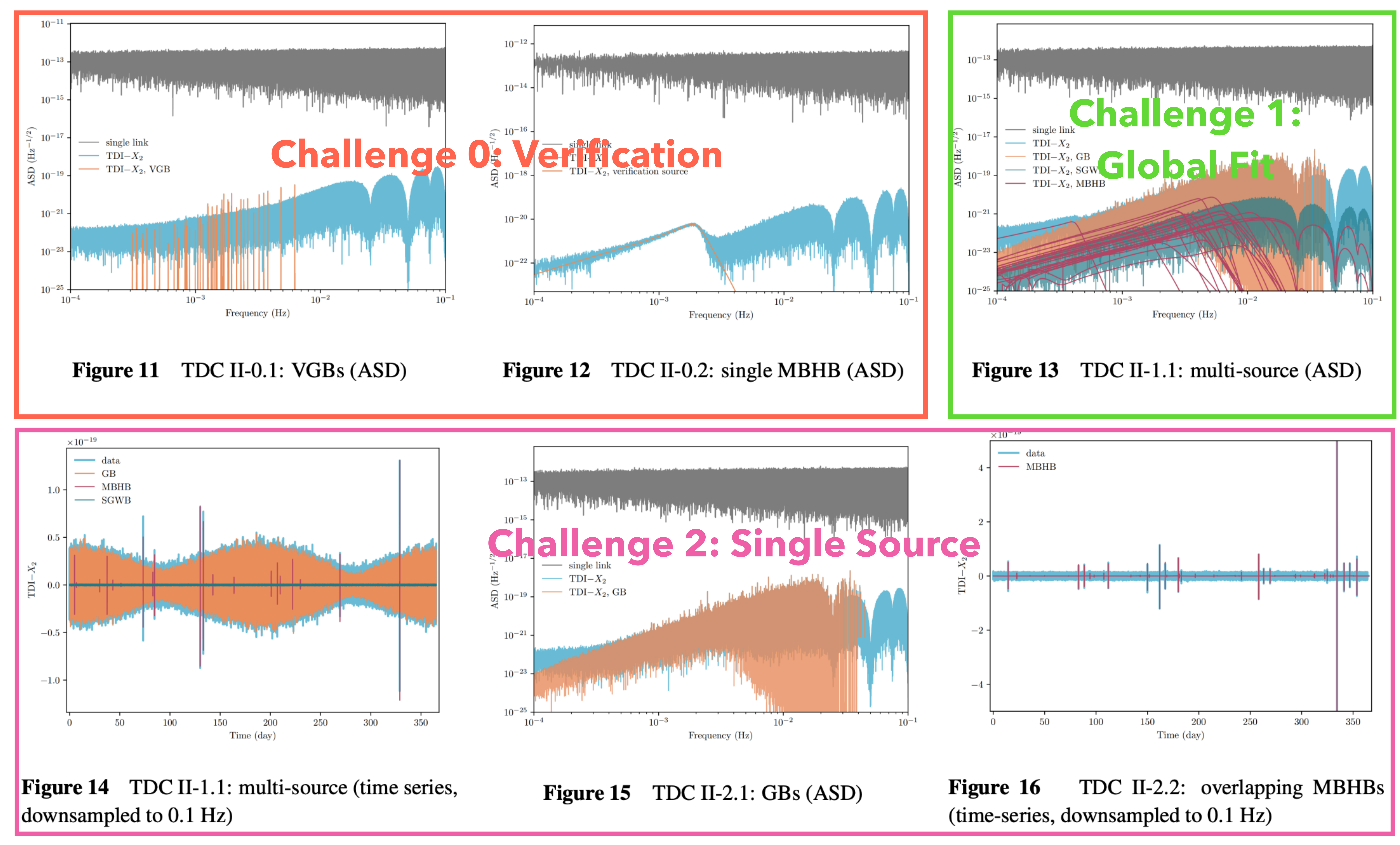
[4] N. Karnesis et al, Phys. Rev. D 104, 043019, 2021

[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

Challenge Datasets



Challenge 0: Verification

Figure 11 TDC II-0.1: VGBs (ASD)

Figure 12 TDC II-0.2: single MBHB (ASD)

Challenge 1: Global Fit

Figure 13 TDC II-1.1: multi-source (ASD)

Challenge 2: Single Source

Figure 14 TDC II-1.1: multi-source (time series, downsampled to 0.1 Hz)

Figure 15 TDC II-2.1: GBs (ASD)

Figure 16 TDC II-2.2: overlapping MBHBs (time-series, downsampled to 0.1 Hz)

Challenge Datasets

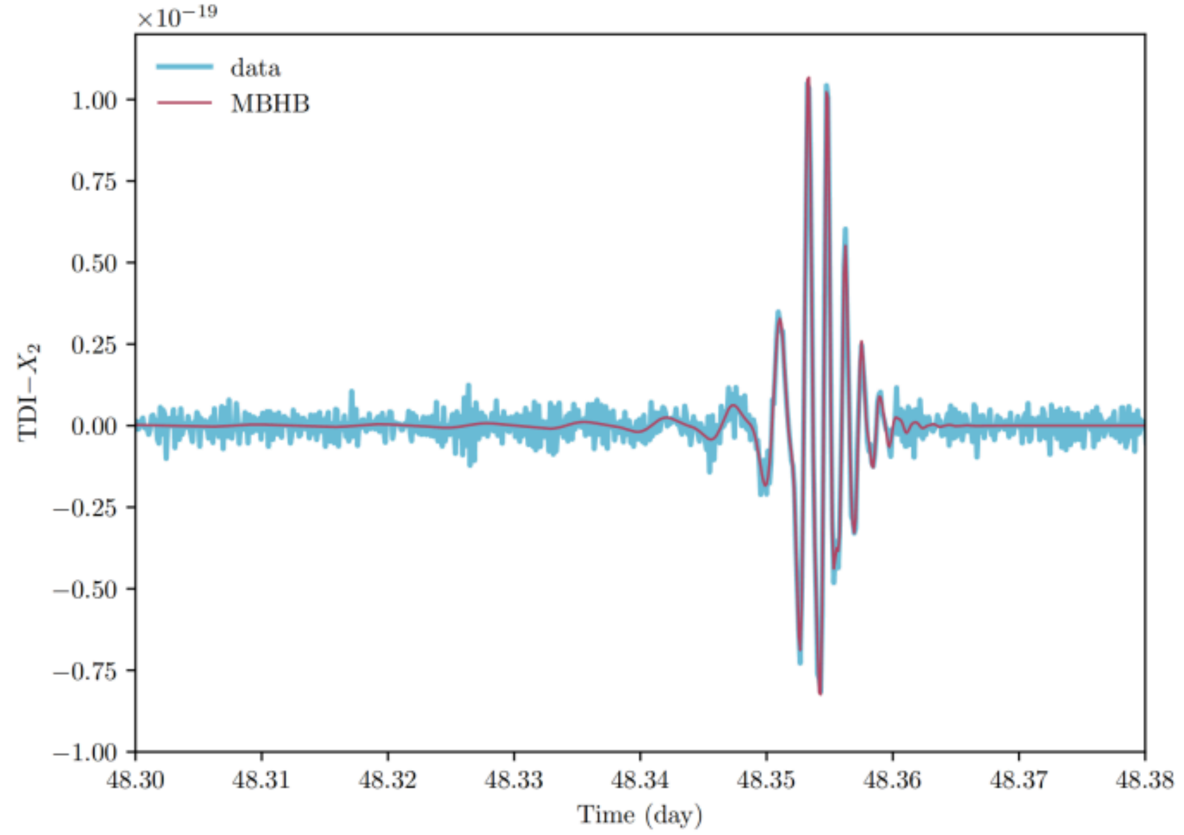


Figure 17 TDC II-2.3: MBHBs with eccentricity and higher harmonics (time-series, downsampled to 0.1 Hz, zoomed in around one of the mergers)

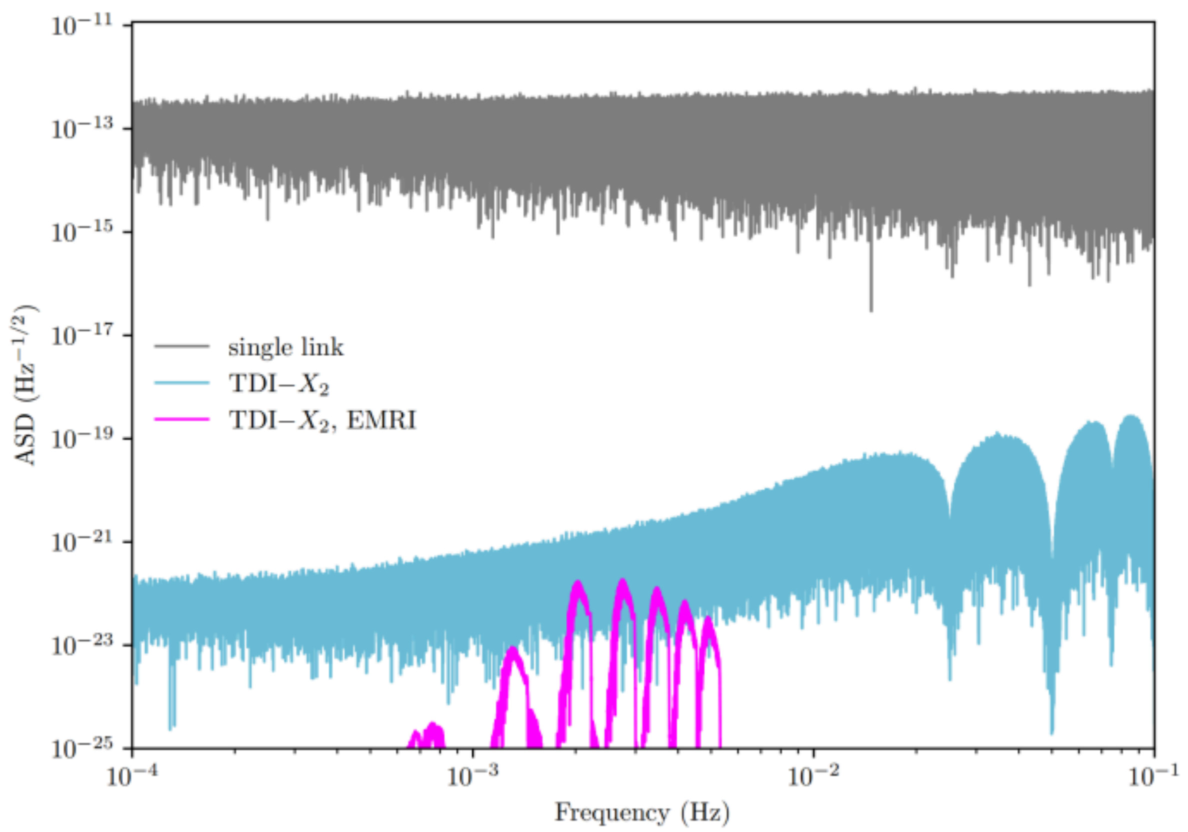


Figure 18 TDC II-2.5: EMRI (ASD, only 1 of the three EMRIs is shown due to their similarity in morphology)

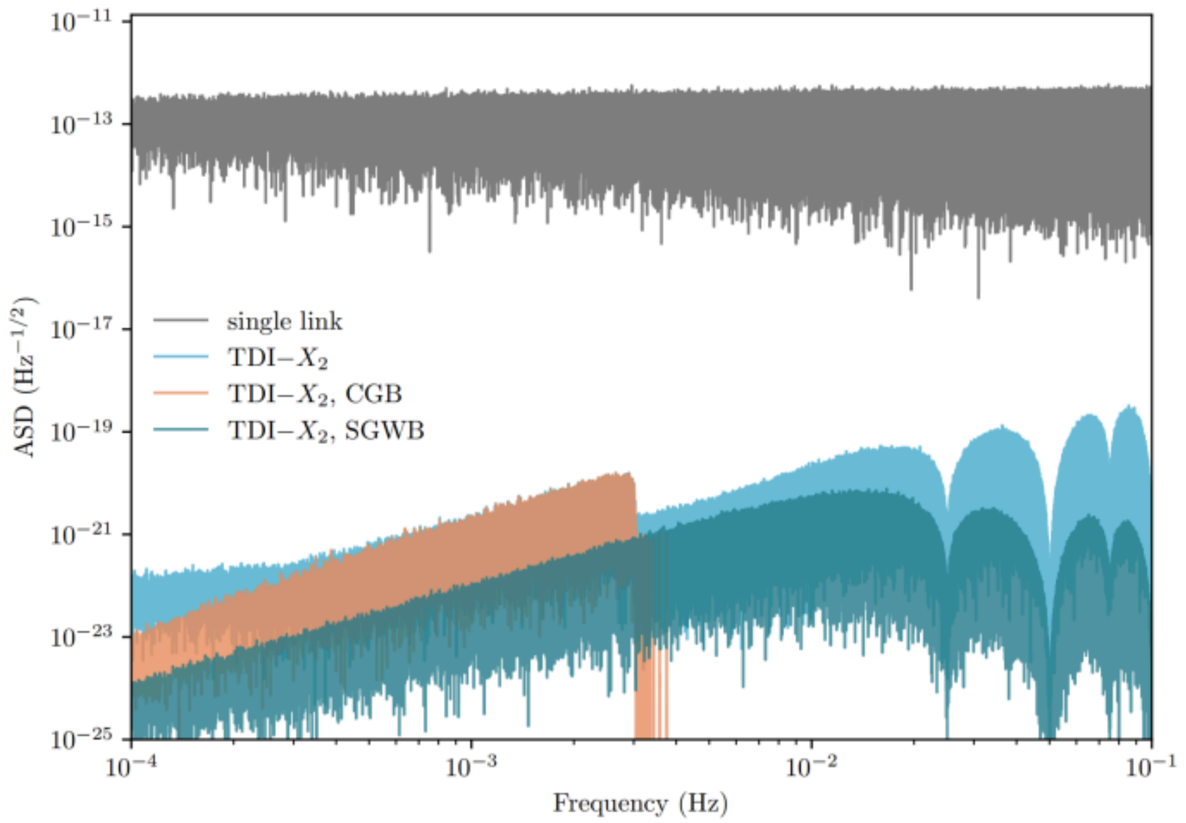


Figure 19 TDC II-2.7: SGWB (ASD)

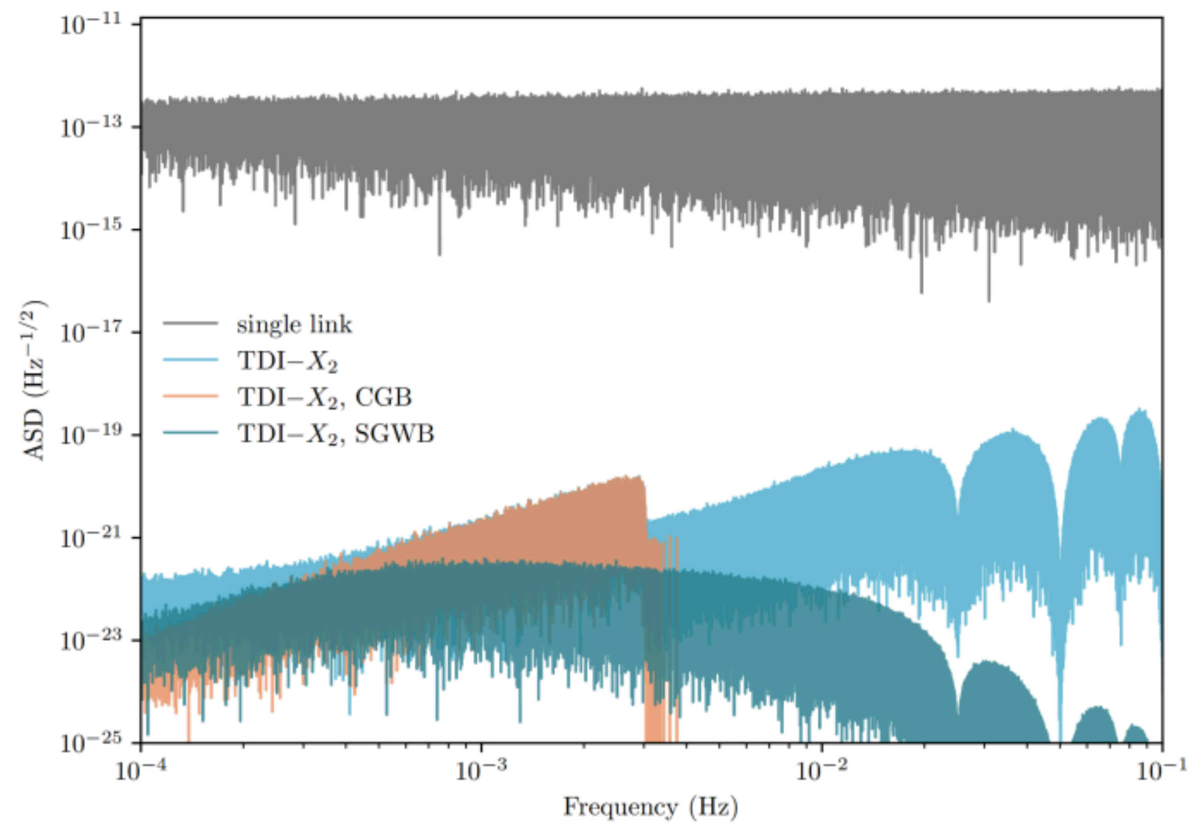


Figure 20 TDC II-2.8: SGWB (ASD, another profile)

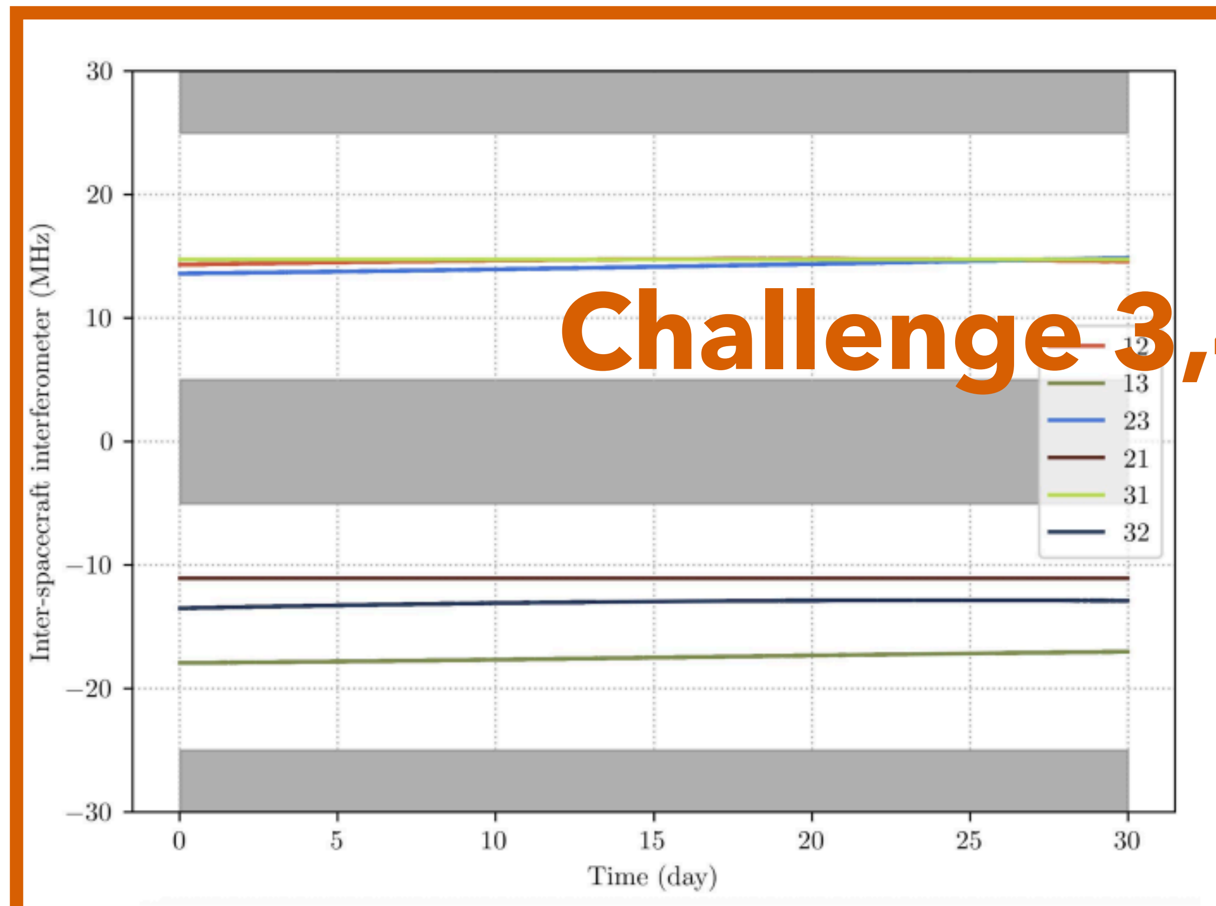


Figure 21 TDC II-3.1: raw ISI readouts (time series)

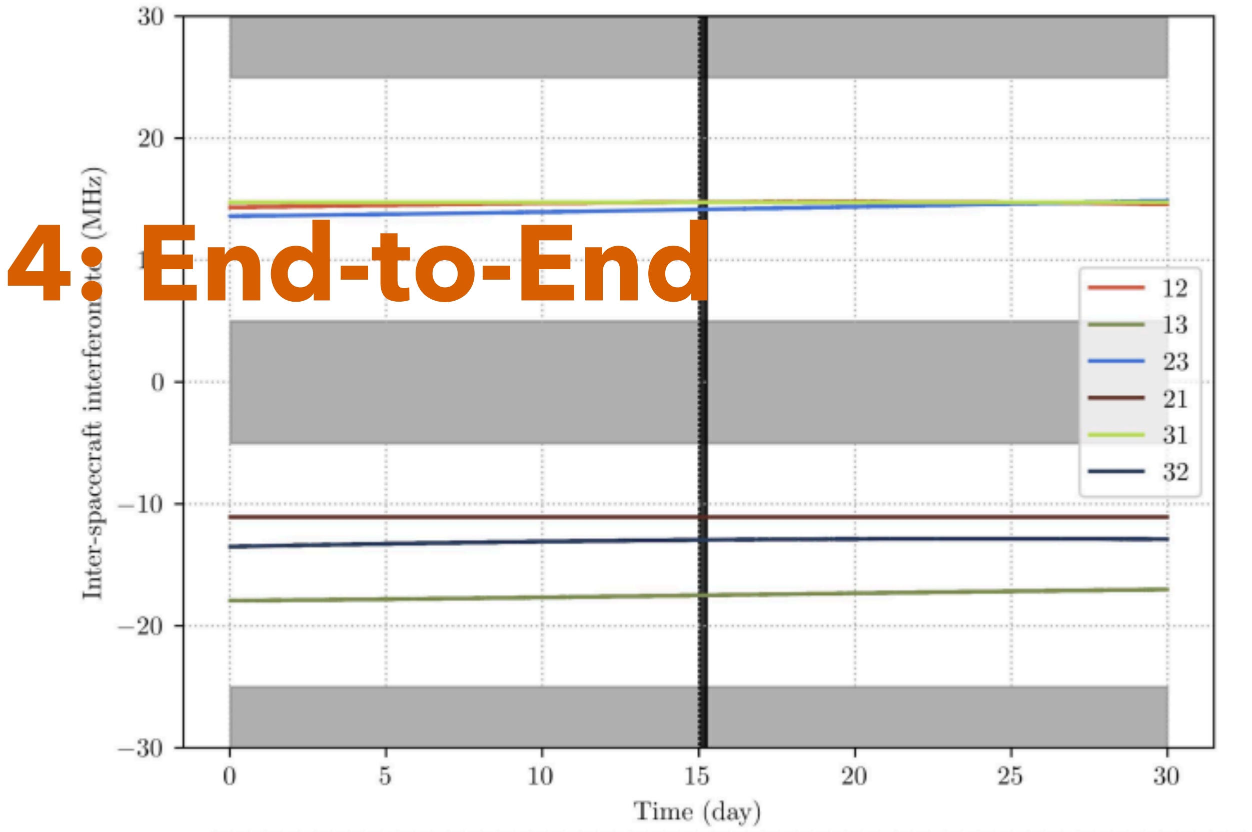
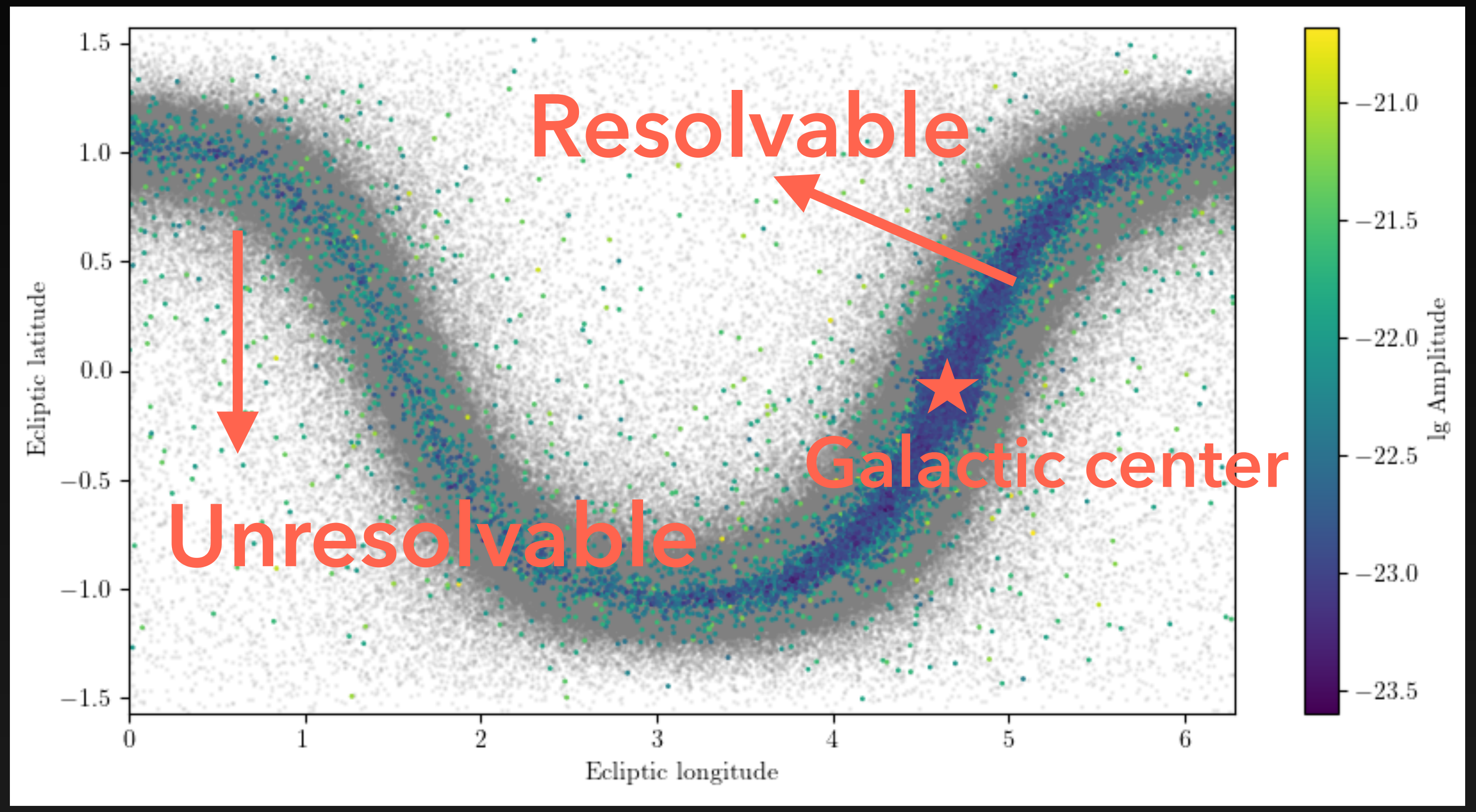


Figure 22 TDC II-4.1: raw ISI readouts (time series, gap shown as a black vertical line)

Challenge 3,4: End-to-End

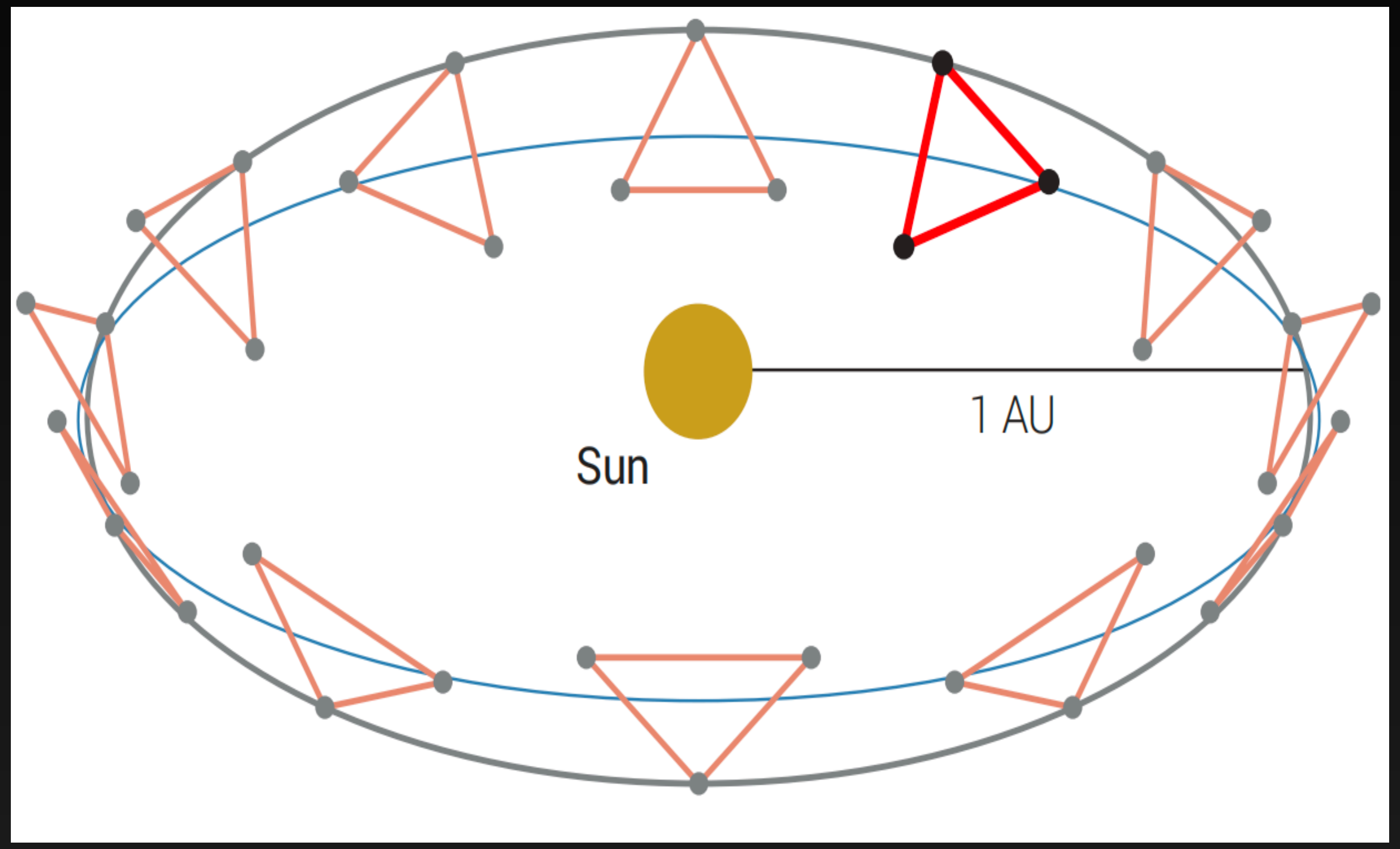
1. Basic concepts for Taiji and space-based GW detection
2. How is Taiji’s data different from LVK
3. The new and more “realistic” Taiji Data Challenge
4. **Non-stationary noises in Taiji’s long-duration observation**
5. Working in the time-frequency domain:
    - Noise characterization
    - Galactic binary estimation
6. Taking-away messages

Non-stationary noises during long-duration observation



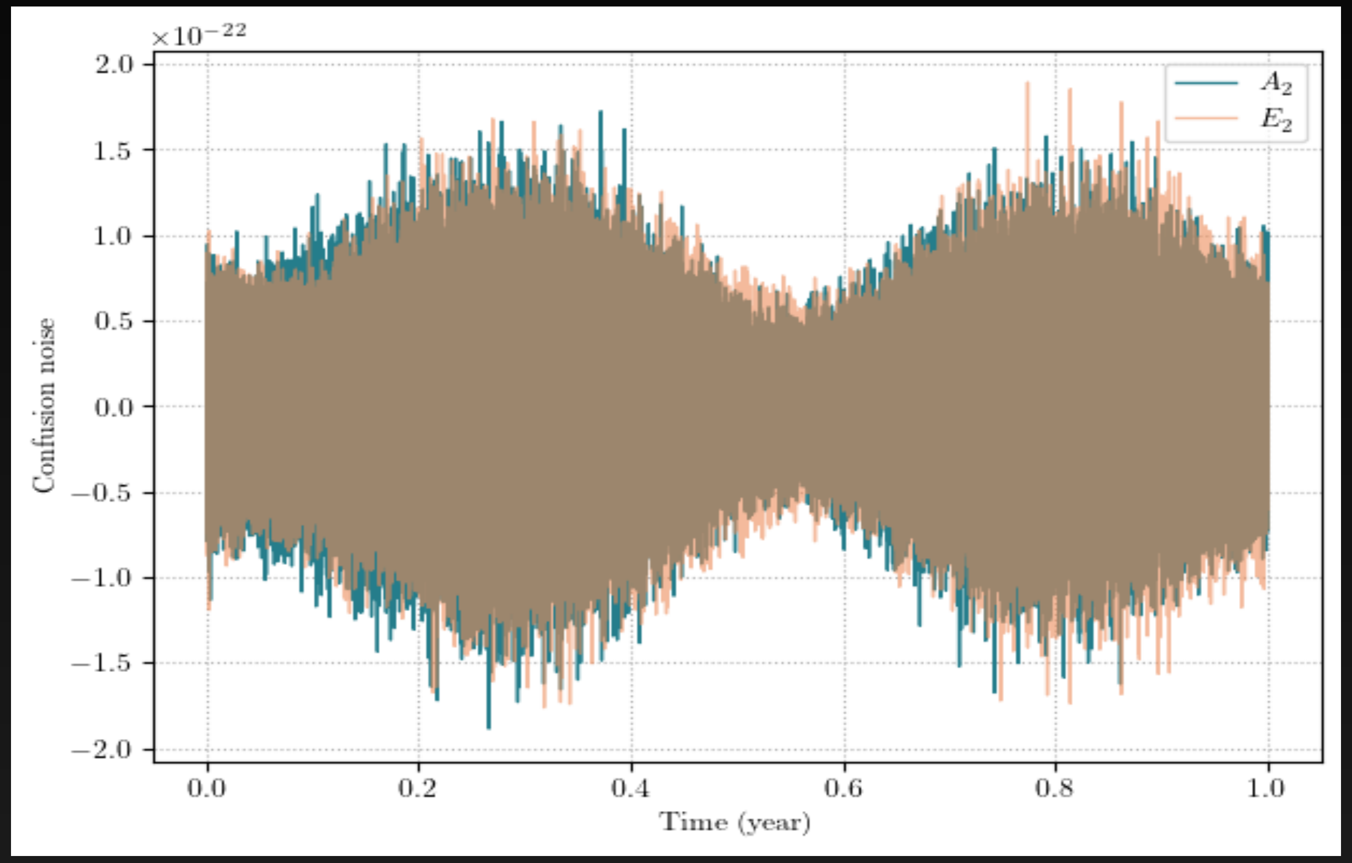
The distribution of GBs
(In Ecliptic coordinate)

+



The rotation of constellation

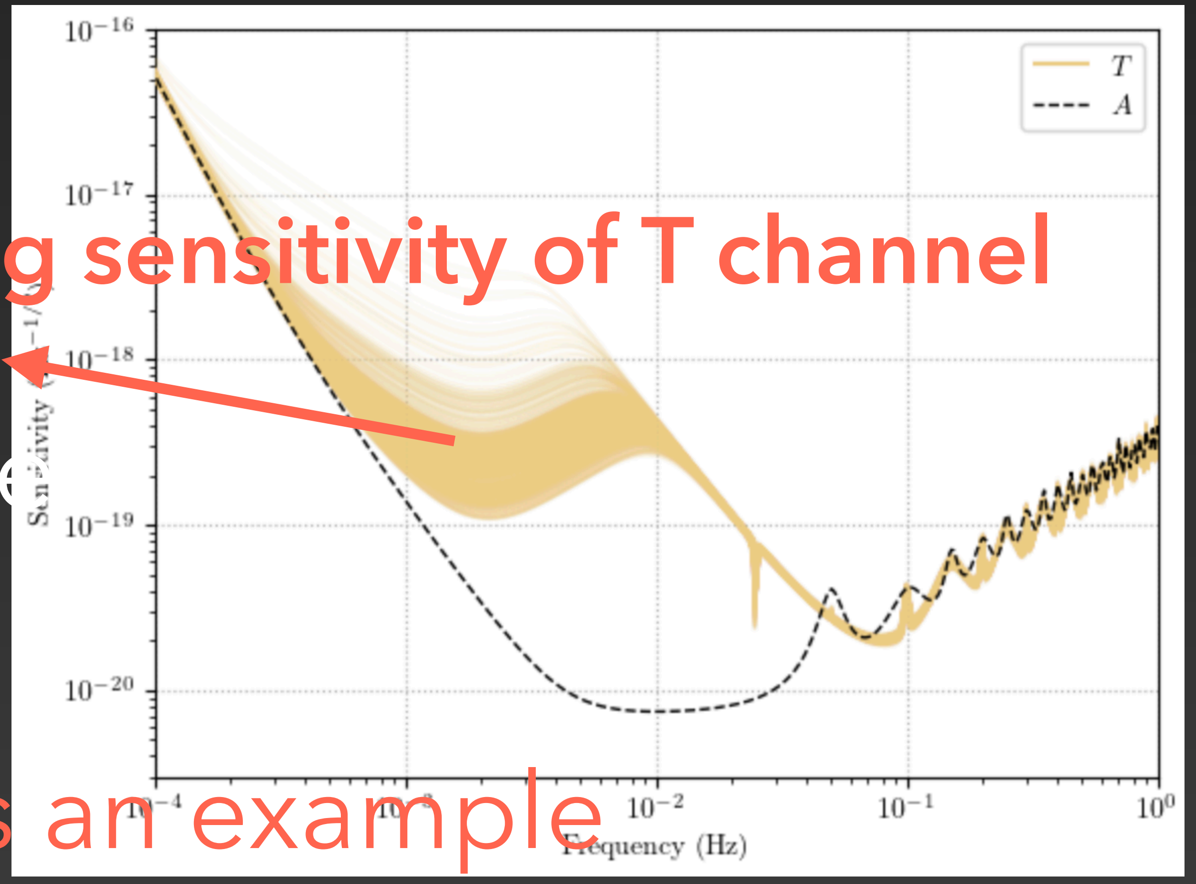
=



The amplitude of confusion foreground vary with time

non-stationary noises might emerge from various origins

- The cyclostationary confusion foreground caused by the anisotropic distribution of GBs and the rotation of constellation
- The drifts of instruments: SC platforms, payloads
- The coupling effect of TDI and arm length variation, e.g. the noise of Michelson T channel is very sensitive to the variation of arms



Time-varying sensitivity of T channel

The case we will use as an example

1. Basic concepts for Taiji and space-based GW detection
2. How is Taiji’s data different from LVK
3. The new and more “realistic” Taiji Data Challenge
4. Non-stationary noises in Taiji’s long-duration observation
5. **Working in the time-frequency domain:**
    - Noise characterization
    - Galactic binary estimation
6. Taking-away messages

2 Data Analysis Tasks Based on Long-Duration Observations

1. Instrumental noise characterization

- Essential for the detection of stochastic GW background (SGWB) and monitoring the detector's operation status
- Incompletely understood at present, obscured by signals in the future
- Using "null" TDI channels (e.g. Michelson-T) sounds ideal, but by parameter degeneracies and non-stationarity (see the following slides)

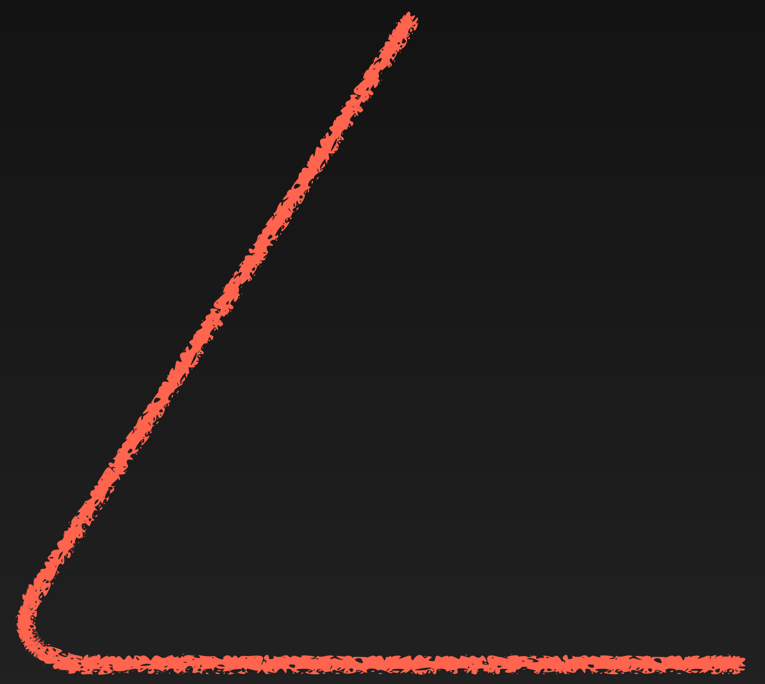
2. Galactic compact binaries (GBs) parameter estimation

- GWs are transparent to interstellar gas, dust, and other stars in the Galaxy -> probes for Galaxy structure and evolution
- $\mathcal{O}(10^4)$ resolvable, $\mathcal{O}(10^7)$ forming a confusion foreground
- Accurate parameter estimation is crucial for both Galactic astrophysics and global fit

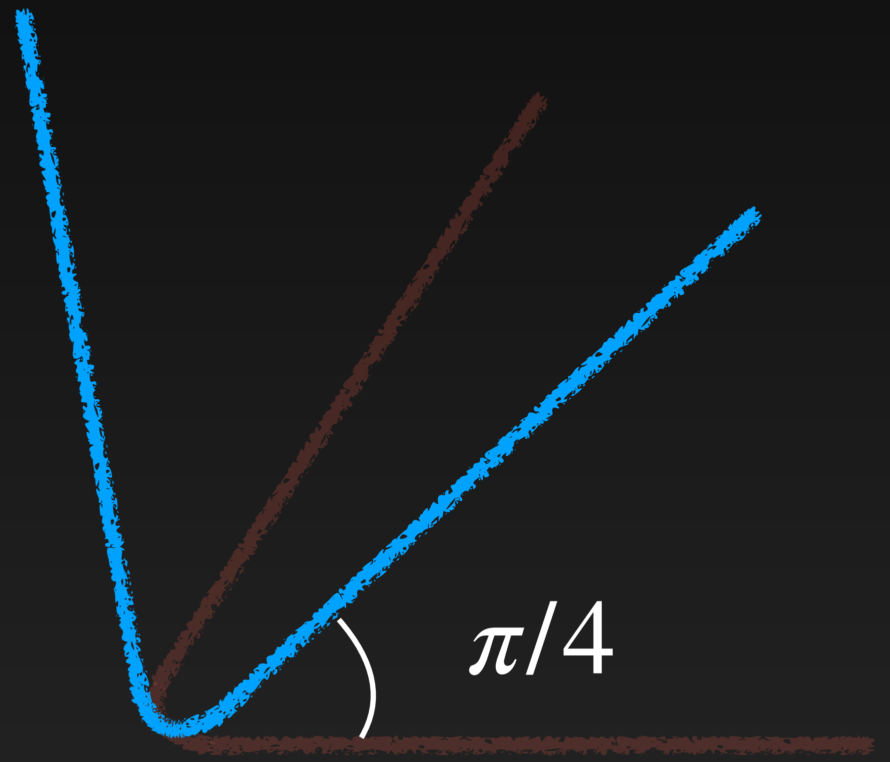
Notes on the T channel

Among all the TDI configurations, the Michelson A / E / T channels are most commonly used.

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



Michelson-A



Michelson-E

Michelson-T: “null”

- **Noise orthogonal:**

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

- **AE signal channels:**

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

- **T insensitive to GWs -> noise channel:**

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

  (Remember that we may not have a lot of extra sensors to monitor the characteristics of noises)

Frequency domain noise estimation

Laser noises & clock noises, suppressed by TDI

- **Single-arm noise model:** $\eta_{ij} = y_{ij} + \boxed{\mathbf{D}_{ij} p_{ji} - p_{ij}} + \boxed{a_{ij} q_i} + \boxed{N_{ij}} + \boxed{\delta_{ij} + \mathbf{D}_{ji} \delta_{ji}}$

OMS noises ACC noises

Noises that will be present in TDI data, to be estimated via T channel

- **TDI noise model:** $S^{\mathrm{TDI}}(f) = \sum_{ij \in \mathrm{MOSAs}} \boxed{\left(A_{ij}^{\mathrm{OMS}}\right)^2} \boxed{F_{ij}^{\mathrm{OMS}}(f)} + \boxed{\left(A_{ij}^{\mathrm{ACC}}\right)^2} \boxed{F_{ij}^{\mathrm{ACC}}(f)}$

$\boldsymbol{\theta}_{\mathrm{noise}} = \{A_{ij}^{\mathrm{OMS}}, A_{ij}^{\mathrm{ACC}}\}$ Noise spectral profiles, assume known

- **Whittle likelihood:** $\ln \mathscr{L}(\boldsymbol{\theta}_{\mathrm{noise}}) = -\frac{1}{2} \sum_f \ln S^{\mathrm{TDI}}(f; \boldsymbol{\theta}_{\mathrm{noise}}) - \frac{1}{2} \sum_f \frac{|\tilde{d}^{\mathrm{TDI}}(f)|^2}{S^{\mathrm{TDI}}(f; \boldsymbol{\theta}_{\mathrm{noise}})}$

Frequency domain noise estimation

$$\eta_{ij} = y_{ij} + \boxed{\mathbf{D}_{ij}p_{ji} - p_{ij}} + \boxed{a_{ij}q_i} + \boxed{N_{ij}} + \boxed{\delta_{ij} + \mathbf{D}_{ji}\delta_{ji}}$$

1. Similar for different spacecrafts, making the parameters highly degenerate

- **TDI noise model:**

$$S^{\mathrm{TDI}}(f) = \sum_{ij \in \mathrm{MOSAs}} \boxed{\left(A_{ij}^{\mathrm{OMS}}\right)^2} \boxed{F_{ij}^{\mathrm{OMS}}(f)} + \boxed{\left(A_{ij}^{\mathrm{ACC}}\right)^2} \boxed{F_{ij}^{\mathrm{ACC}}(f)}$$

$$\boldsymbol{\theta}_{\mathrm{noise}} = \{A_{ij}^{\mathrm{OMS}}, A_{ij}^{\mathrm{ACC}}\}$$

$$\ln \mathscr{L}(\boldsymbol{\theta}_{\mathrm{noise}}) = -\frac{1}{2}\sum_{f} \ln S^{\mathrm{TDI}}(f; \boldsymbol{\theta}_{\mathrm{noise}}) - \frac{1}{2}\sum_{f} \frac{|\tilde{d}^{\mathrm{TDI}}(f)|^2}{S^{\mathrm{TDI}}(f; \boldsymbol{\theta}_{\mathrm{noise}})}$$

Frequency domain noise estimation

$$\eta_{ij} = y_{ij} + \boxed{\mathbf{D}_{ij}p_{ji} - p_{ij}} + \boxed{a_{ij}q_i} + \boxed{N_{ij}} + \underbrace{\boxed{\delta_{ij} + \mathbf{D}_{ji}\delta_{ji}}}_{}$$

$$S^{\mathrm{TDI}}(f) = \sum_{ij \in \mathrm{MOSA}} \boxed{\left(A_{ij}^{\mathrm{OMS}}\right)^2} \boxed{F_{ij}^{\mathrm{OMS}}(f)} + \boxed{\left(A_{ij}^{\mathrm{ACC}}\right)^2} \boxed{F_{ij}^{\mathrm{ACC}}(f)}$$

2. The non-stationary of T channel would violate the fundamental assumption of Whittle likelihood. In this case, different frequencies will be correlated. Rigorous calculation of L would become computational intensive.

$$\boldsymbol{\theta}_{\mathrm{noise}} = \{A_{ij}^{\mathrm{OMS}}, A_{ij}^{\mathrm{ACC}}\}$$

- **Whittle likelihood:**

$$\ln \mathscr{L}(\boldsymbol{\theta}_{\mathrm{noise}}) = -\frac{1}{2}\sum_f \ln S^{\mathrm{TDI}}(f;\boldsymbol{\theta}_{\mathrm{noise}}) - \frac{1}{2}\sum_f \frac{|\tilde{d}^{\mathrm{TDI}}(f)|^2}{S^{\mathrm{TDI}}(f;\boldsymbol{\theta}_{\mathrm{noise}})}$$

Time-frequency domain noise estimation

(WDM wavelet basis)

A natural modification is to move to the time-frequency domain

The time-dependence is due to arm length variation, which can be measured to a high precision,
So these F(t, f) functions can be well modeled

- **TDI noise model:**

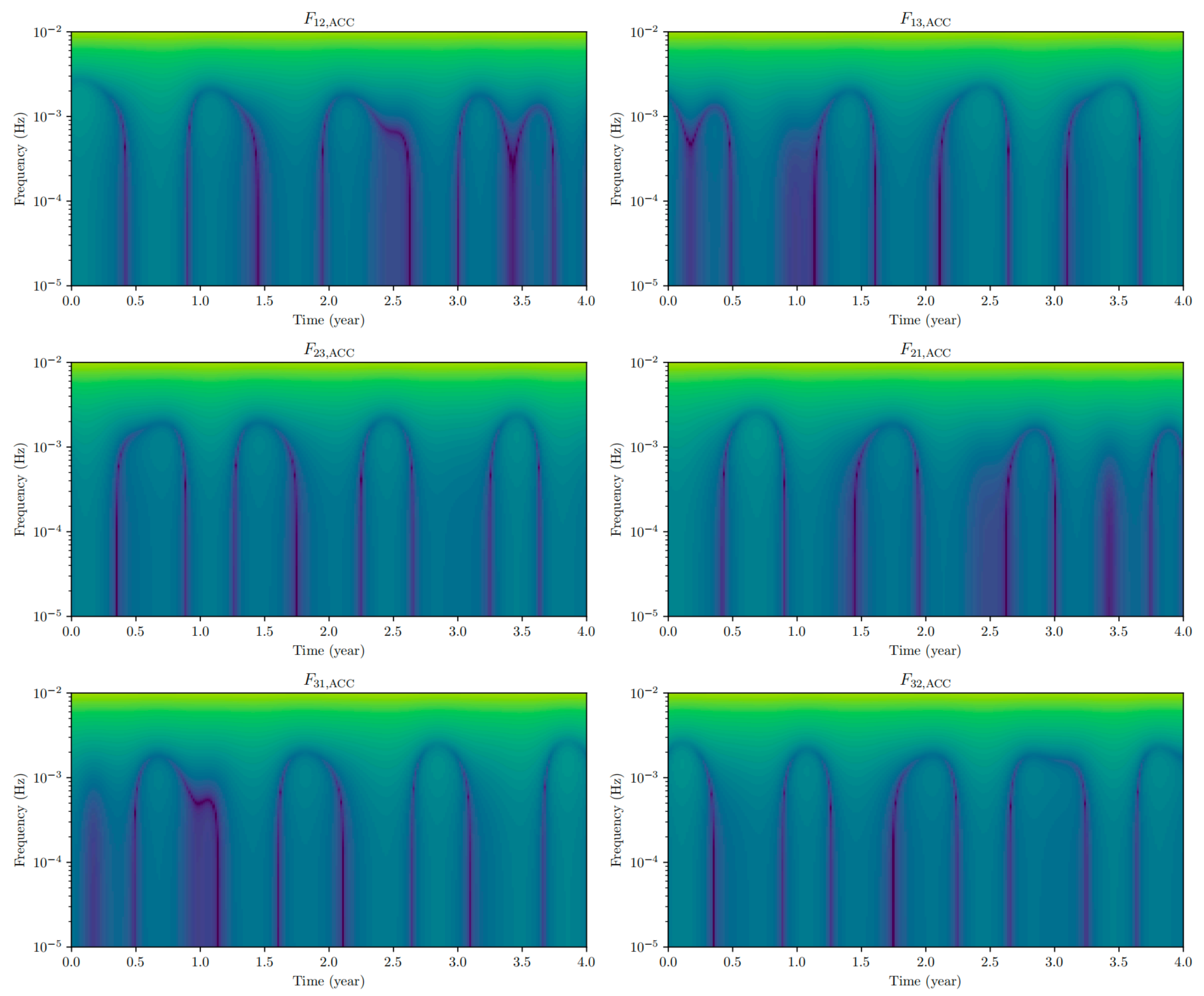
$$S_{nm}^{\mathrm{TDI}}(\boldsymbol{\theta}_{\mathrm{noise}}) = \sum_{ij \in \mathrm{MOSAs}} \left(A_{ij}^{\mathrm{OMS}}\right)^2 F_{ij}^{\mathrm{OMS}}(\boxed{t_n}, f_m) + \left(A_{ij}^{\mathrm{ACC}}\right)^2 F_{ij}^{\mathrm{ACC}}(\boxed{t_n}, f_m)$$

- **likelihood:**

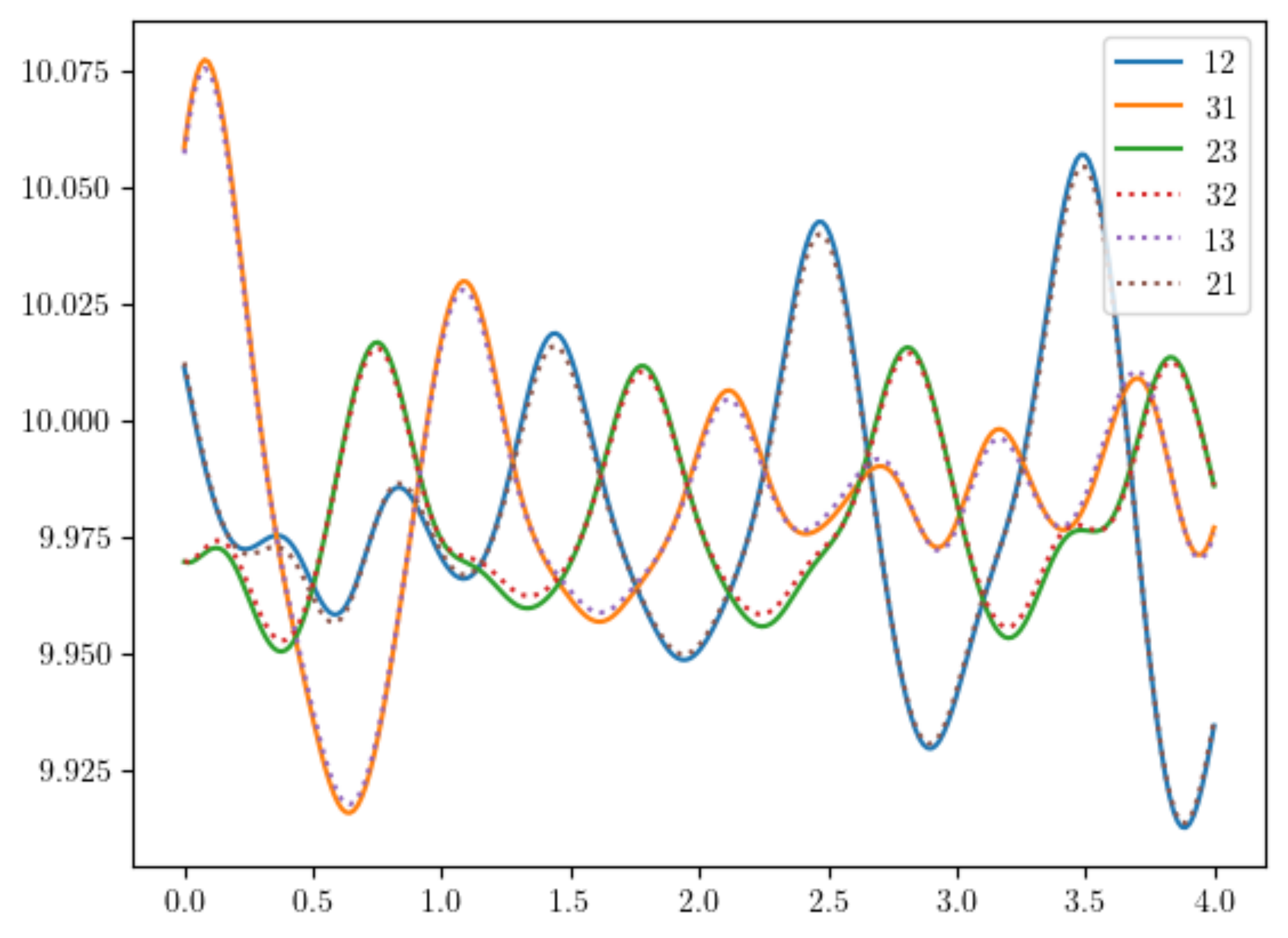
$$\boxed{\ln \mathscr{L}(\boldsymbol{\theta}_{\mathrm{noise}}) = -\frac{1}{2} \sum_{nm} \ln S_{nm}^{\mathrm{TDI}}(\boldsymbol{\theta}_{\mathrm{noise}}) - \frac{1}{2} \sum_{f} \frac{|\tilde{d}_{nm}^{\mathrm{TDI}}|^2}{S_{nm}^{\mathrm{TDI}}(\boldsymbol{\theta}_{\mathrm{noise}})}}$$

Based on local stationary condition (the arms are slow varying), all the time-frequency pixels are independent, so the likelihood can still quickly calculated

The time-frequency spectral shape $F_{ij}^{\mathrm{ACC}}(t_n, f_m)$



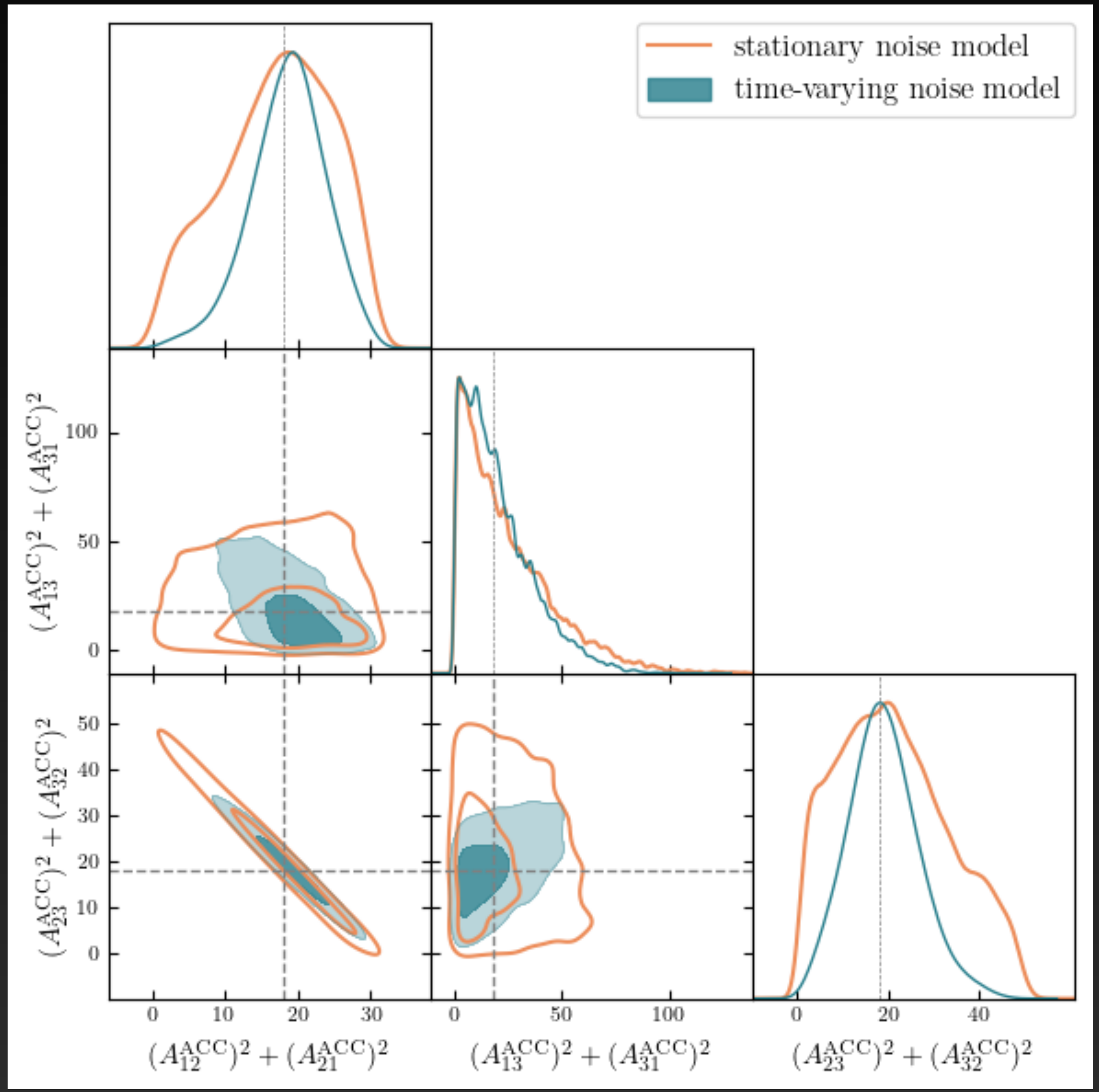
$F_{12,\mathrm{ACC}}$ $F_{13,\mathrm{ACC}}$ $F_{23,\mathrm{ACC}}$ $F_{21,\mathrm{ACC}}$ $F_{31,\mathrm{ACC}}$ $F_{32,\mathrm{ACC}}$



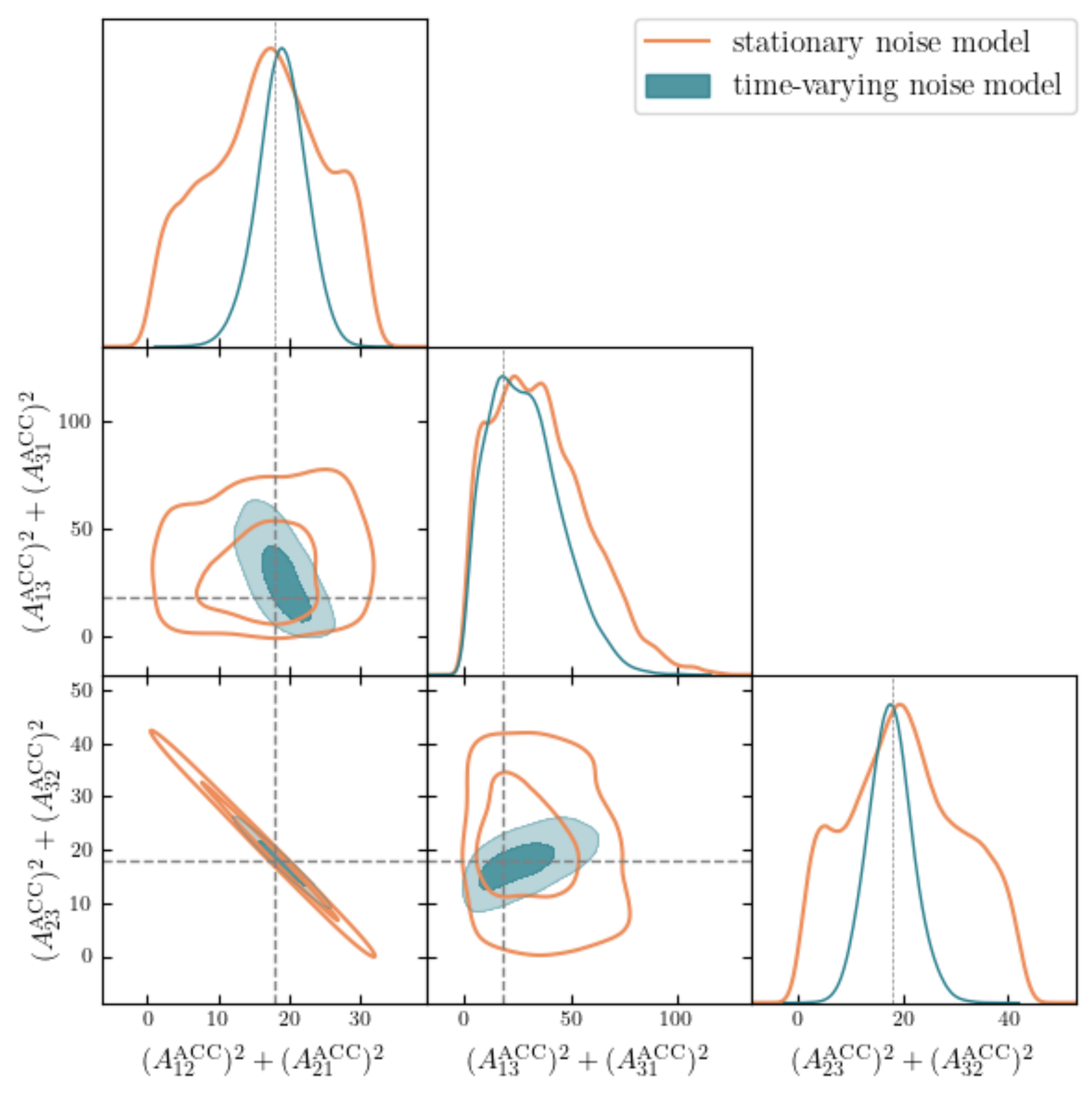
Arm length variation

We can actually leverage the variation of arm to break the degeneracy among parameters, since when considering the time-dependence, the contributions of noise parameters are now distinguishable.

Noise estimation of T channel



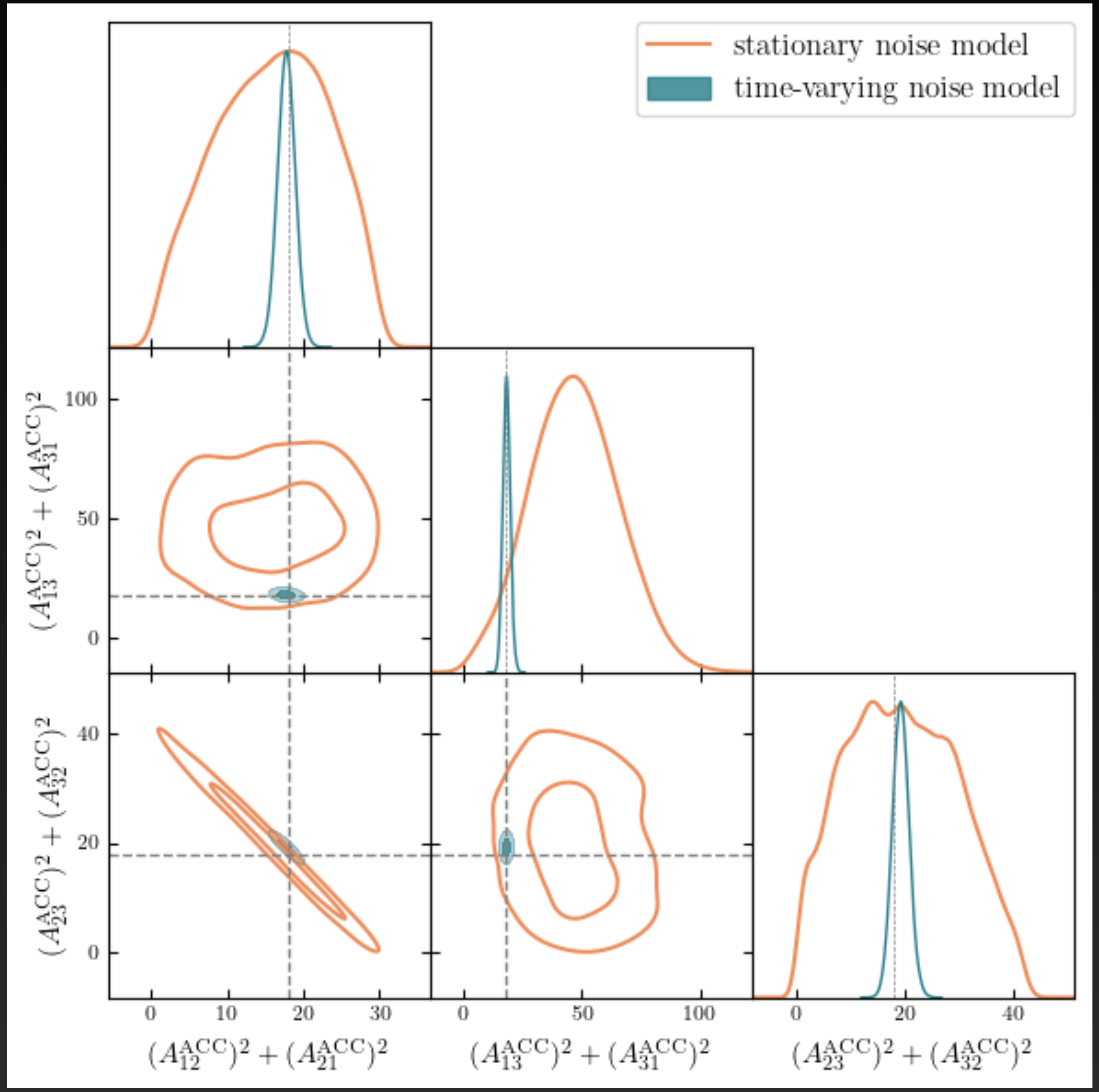
1 month



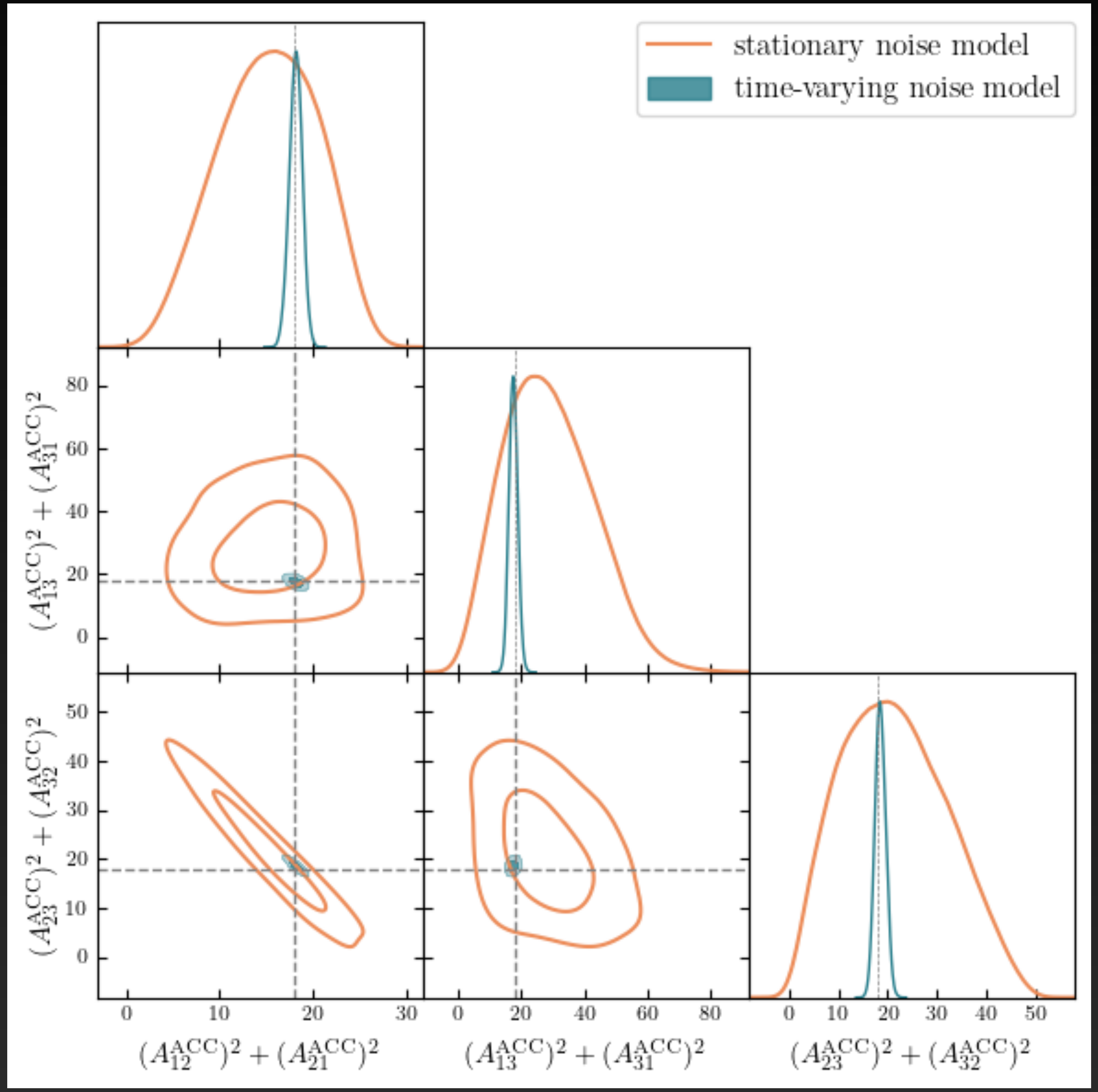
1/4 year

using the this formalism, the parameters can be better constrained, and the effect gets better for longer observational times

Noise estimation of T channel



1/2 year



1 year

The parameters of time-varying noises can be estimated as long as they can be well modeled. With similar method, Phys. Rev. D 111, 023025 (2025) provides an example on the confusion foreground of LISA

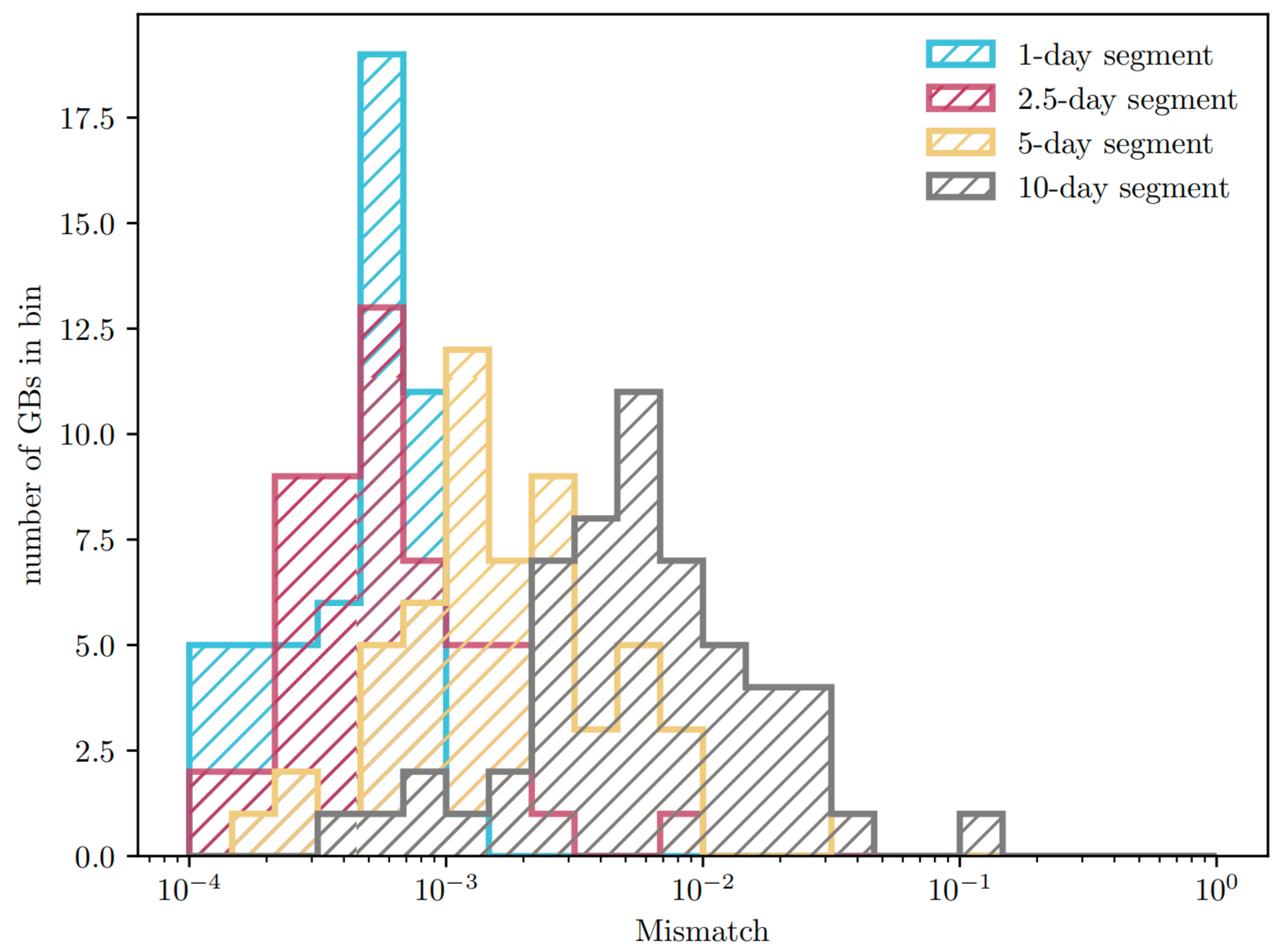
Short-time frequency transform GB template

$$\bar{X}_2(t_m, f) \equiv \int_0^{T_m} \mathrm{d}t\; e^{-i2\pi f t} w_m\left(t - \frac{T_m}{2}\right) \times\; X_2\left(t + t_m - \frac{T}{2}\right)$$

$$\approx e^{-i\pi f T}\tilde{w}\left[f - \frac{\dot{\varphi}(t_m)}{2\pi}\right] A_{X_2}(t_m) \qquad \Delta\Phi < 1/\mathrm{SNR}$$

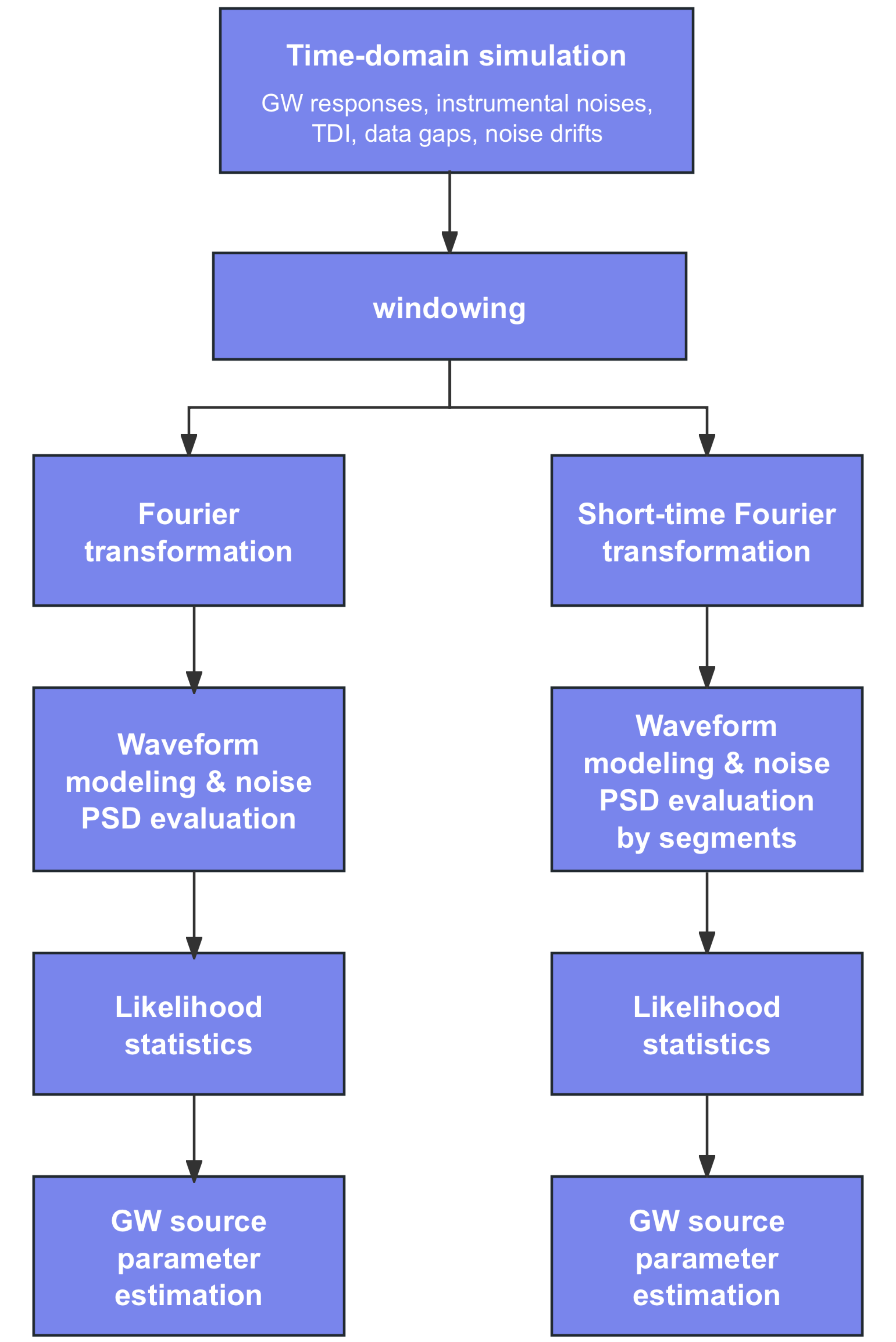
$$A_{X_2}(t_m) = \Big\{ \left(1 - \Delta_{31}^2\right) \left[B_{y,12}(t_m) + \Delta_{12} B_{y,21}(t_m)\right] - \left(1 - \Delta_{12}^2\right) \left[B_{y,13}(t_m) + \Delta_{31} B_{y,31}(t_m)\right] \Big\} \times \left(1 - \Delta_{12}^2 \Delta_{31}^2\right) e^{i\varphi(t_m)}$$

$$\langle d\,|\,h\rangle \equiv 4 \sum_{t_m}^{Nt} \frac{1}{T_m} \sum_{f_{m,n}}^{N_{f,m}} \frac{\Re\left[\bar{d}\left(t_m, f_{m,n}\right) \bar{h}^*\left(t_m, f_{m,n}\right)\right]}{S\left(t_m, f_{m,n}\right)}$$



Mismatches of STFT template and time-domain simulation (using TDC toolkit)

The deduction ensures $\Delta\Phi < 1/\mathrm{SNR}$, which is a threshold derived from the requirement of unbiased estimation [Lindblom, 0809.3844]. The same threshold is also seen in the literature of EMRI waveforms.



Time-domain simulation

GW responses, instrumental noises, TDI, data gaps, noise drifts

windowing

Fourier transformation

Waveform modeling & noise PSD evaluation

Likelihood statistics

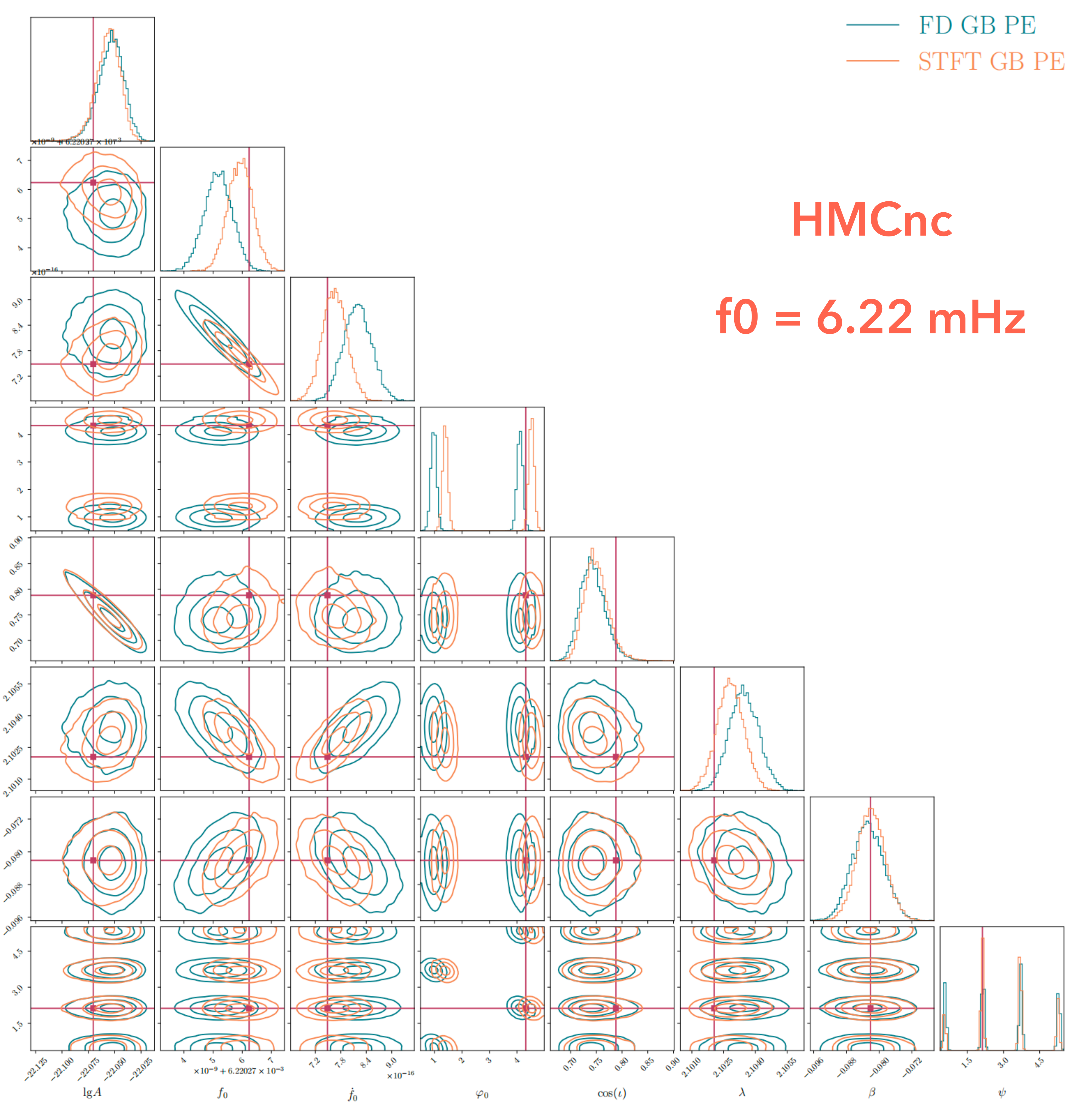
GW source parameter estimation

Short-time Fourier transformation

Waveform modeling & noise PSD evaluation by segments

Likelihood statistics

GW source parameter estimation

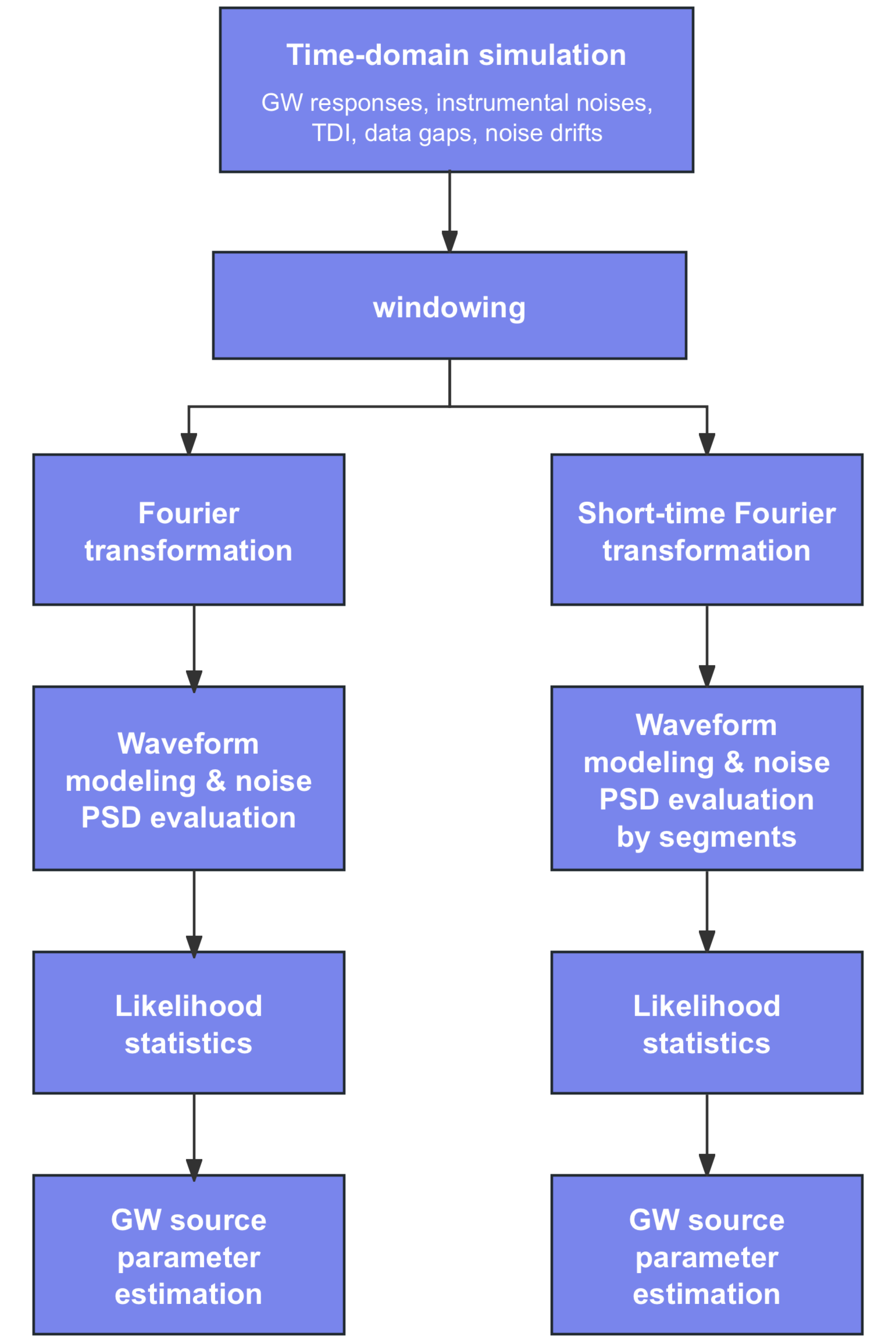


HMCnc

f0 = 6.22 mHz

Model signal & noise in FD / STFT domain

GB posteriors using FD / STFT domain estimation



Time-domain simulation

GW responses, instrumental noises, TDI, data gaps, noise drifts

windowing

Fourier transformation

Waveform modeling & noise PSD evaluation

Likelihood statistics

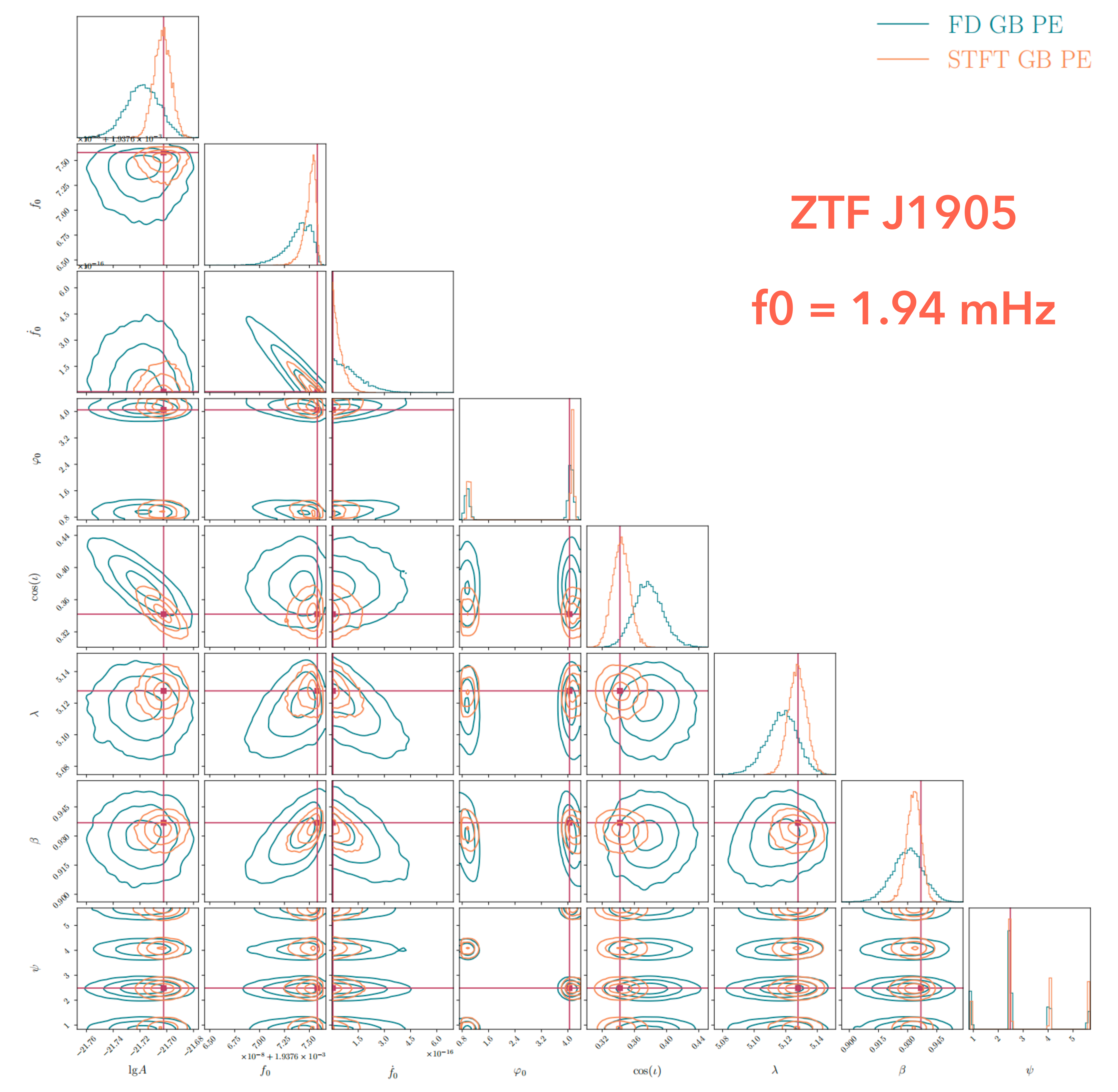
GW source parameter estimation

Short-time Fourier transformation

Waveform modeling & noise PSD evaluation by segments

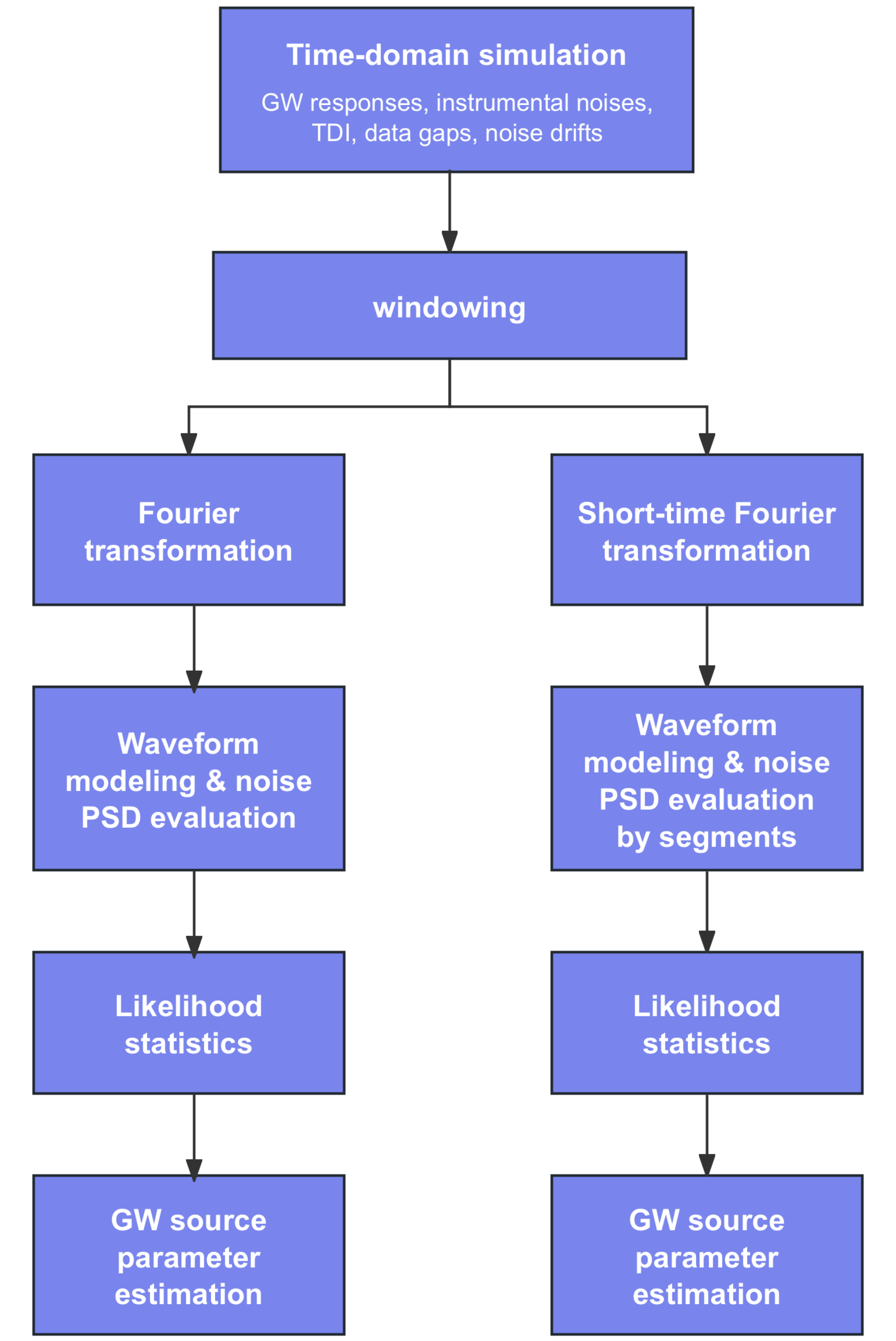
Likelihood statistics

GW source parameter estimation



Model signal & noise in FD / STFT domain

GB posteriors using FD / STFT domain estimation



Time-domain simulation

GW responses, instrumental noises, TDI, data gaps, noise drifts

windowing

Fourier transformation

Waveform modeling & noise PSD evaluation

Likelihood statistics

GW source parameter estimation

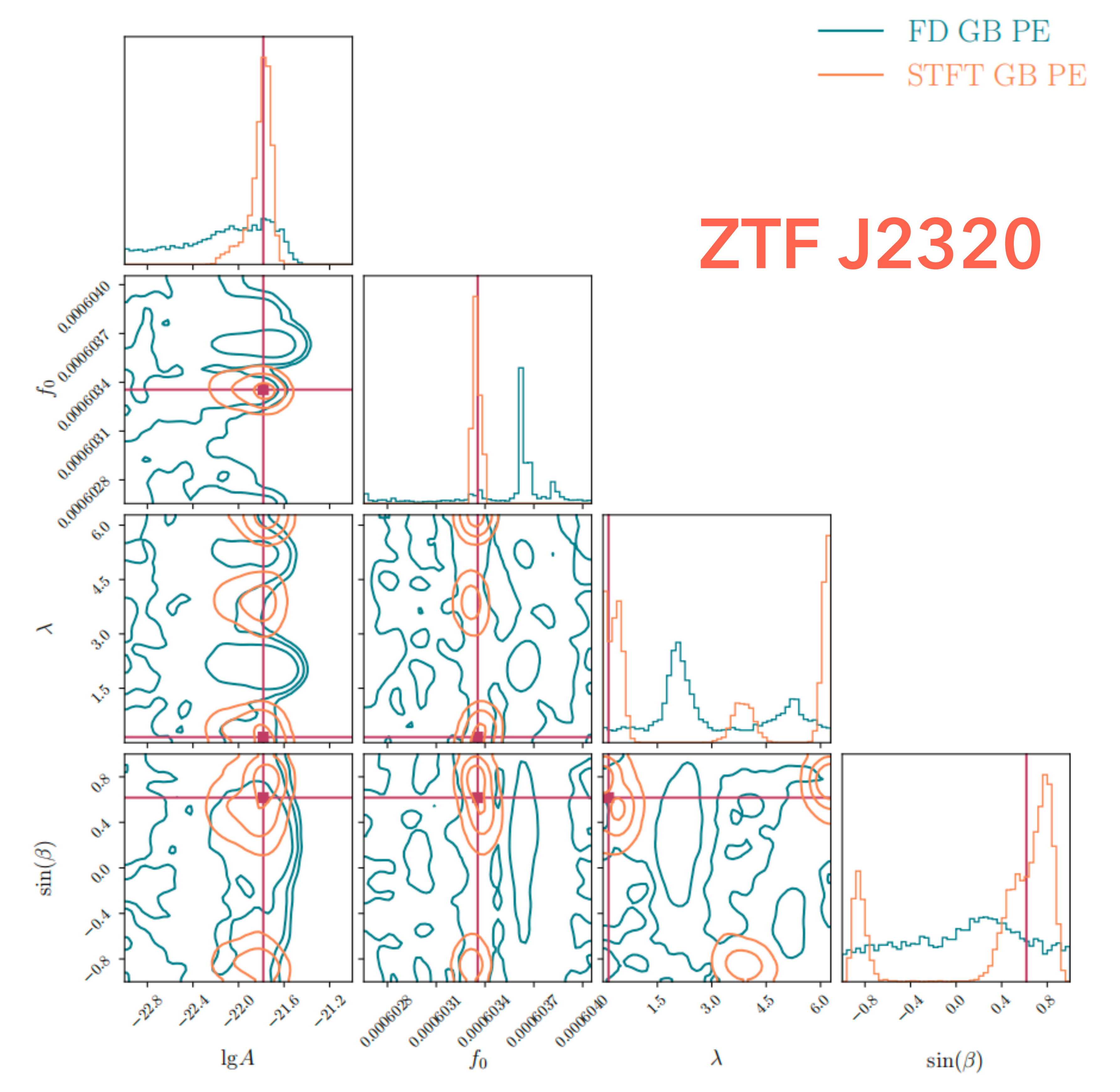
Short-time Fourier transformation

Waveform modeling & noise PSD evaluation by segments

Likelihood statistics

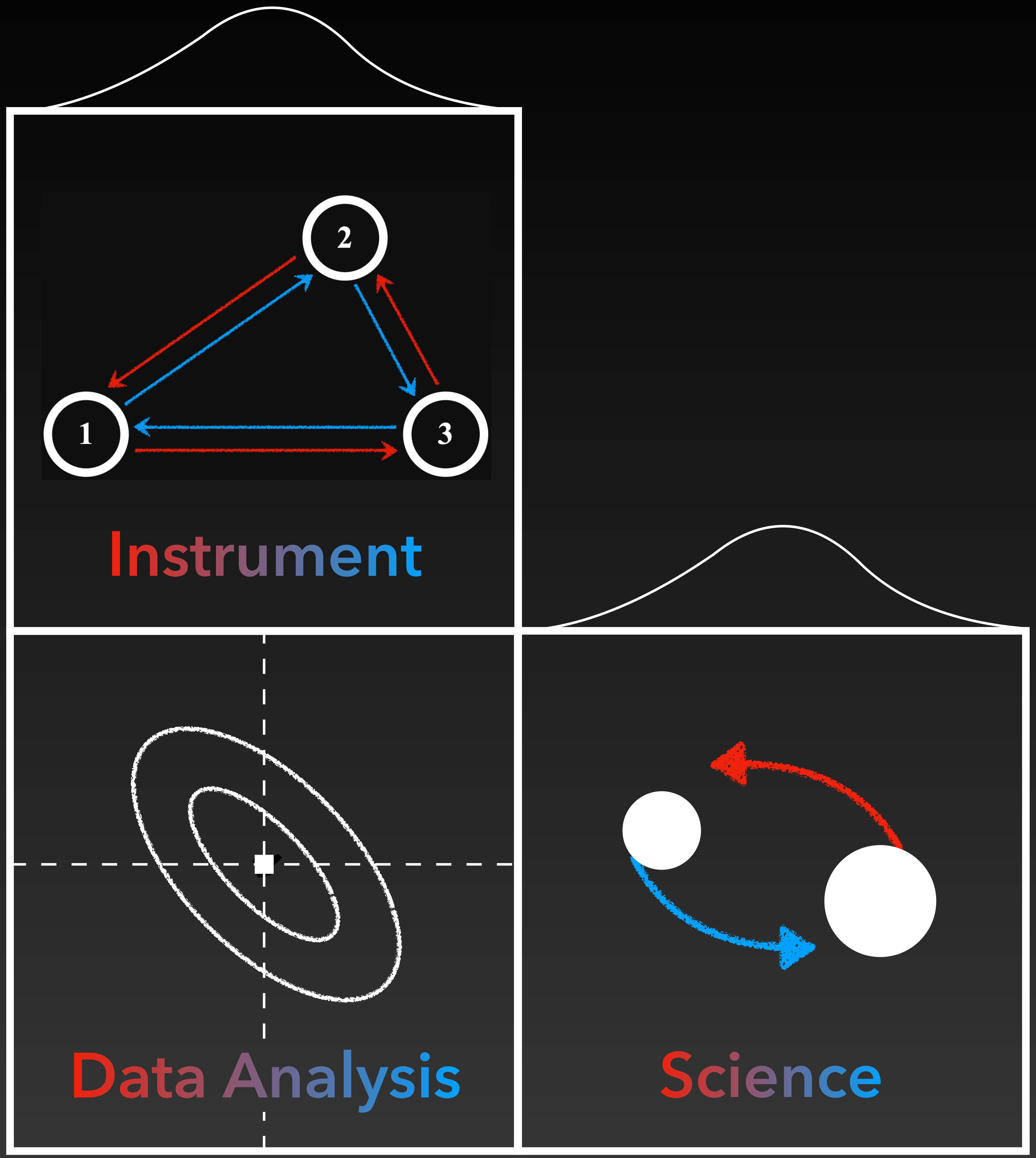
GW source parameter estimation

Model signal & noise in FD / STFT domain



GB posteriors using FD / STFT domain estimation

1. Basic concepts for Taiji and space-based GW detection
2. How is Taiji's data different from LVK
3. The new and more "realistic" Taiji Data Challenge
4. Non-stationary noises in Taiji's long-duration observation
5. Working in the time-frequency domain:
    - Noise characterization
    - Galactic binary estimation
6. **Taking-away messages**



The TDC working group

Taiji Data Challenge

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

Available after May, 23, 2025

Website: http://gr.imech.ac.cn/overview/

GitHub: https://github.com/TriangleDataCenter

- Download links of Data
- Manual, Presentation
- Tools and Tutorials