

The impact of missing data on the construction of LISA Time Delay Interferometry Michelson variables

Ollie Burke^{1,*}, Martina Muratore^{2,†} and Graham Woan¹

¹*School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, UK*

²*Max Planck Institute for Gravitational Physics (Albert Einstein Institute), D-14476 Potsdam, Germany*

(Dated: October 9, 2025)

We investigate the impact of missing input data on the construction of second-generation Time Delay Interferometry (TDI) variables, which enable data analysis for the Laser Interferometer Space Antenna (LISA). TDI relies on the introduction of precise time delays into the raw interferometric data streams before they are combined to suppress otherwise dominant laser phase noise. We show that a single missing sample, corresponding to 0.25 s of data, will result in an effective data gap of approximately 90 s in the second-generation TDI output if further measures are not taken. This additional gap is largely independent of the initial gap duration, but increases linearly with the order of the fractional-delay filter used for the computations. For a realistic gap scenario, incorporating both planned and unplanned data interruptions consistent with a target duty cycle of $\sim 84\%$, we find that frequent, short-duration gaps (e.g., a total of 1000 per year, each of which have short durations ~ 100 s) could result in an additional loss in the TDI variables of about one day per year corresponding to a $\sim 0.3\%$ reduction in duty cycle. This amounts to a loss of approximately one day of LISA data suitable for the global-fit per year.

I. INTRODUCTION

The Laser Interferometer Space Antenna (LISA) is a European Space Agency led space-based gravitational wave (GW) observatory with the goal of observing astrophysical gravitational waves frequencies between 0.1 mHz and 1 Hz [1, 2]. Unlike ground-based detectors, LISA is expected to be signal dominated, observing a multitude of GWs that overlap in both time and frequency [3, 4]. Current GW inference techniques for LISA are not designed to operate directly on the raw phasemeter data. Instead, they take pre-processed data from an initial noise reduction pipeline that transforms the raw phasemeter measurements into calibrated data streams suitable for analysis [5]. This is due to the dominance of laser frequency noise—arising both from intrinsic laser instability and arm-length variations—which exceeds the (GW) astrophysical signal by eight to nine orders of magnitude in amplitude. In contrast, ground-based GW detectors like LIGO, VIRGO and KAGRA are able to suppress laser noise almost perfectly by constructing near-equal arm interferometers [6, 7]. LISA operates as a formation-flying constellation in space, with three spacecraft arranged in an approximately equilateral triangle that trails Earth in a heliocentric orbit inclined by 60 degrees to the ecliptic [8]. While the nominal arm lengths are $\sim 2.5 \times 10^6$ km, they are neither equal nor constant over the mission lifetime. As a result, LISA cannot adopt a traditional equal-arm Michelson interferometer design. To address this, LISA

employs Time-Delay Interferometry (TDI), a technique that synthesizes virtual equal-arm interferometers by appropriately time-shifting and linearly combining measurements taken along different arms. This process effectively suppresses the laser frequency noise to levels compatible with GW science objectives [9–14].

Each LISA spacecraft (S/C) hosts two Moving Optical Sub Assemblies (MOSAs), and each of these carries an optical bench equipped with three distinct interferometers (IFOs): (i) an inter-satellite IFO (ISI), (ii) a test-mass IFO (TMI), and (iii) a reference IFO (RFI) [1]. The test-mass IFO measures sub-nanometre fluctuations in the displacement between the optical bench (OB) and its associated free-falling test mass which provides a local inertial reference. The inter-satellite IFO interferes the incoming light from a remote spacecraft with locally generated laser light, measuring the relative optical path length between distant OB and local OB. Finally, the reference interferometer measures the relative displacement between the two OBs hosted on the same S/C. These give a total of 18 interferometric phasemeter measurements. In total however, the full raw phasemeter data has 66 channels, including clock sidebands and measured pseudoranges (MPRS) comprising the inter-spacecraft distance plus clock offset and ranging noises, that are needed to fully correct for timing and laser noise. Additional differential wavefront sensors measure the laser beam angles, and are used to minimize the effects of angular jitter [15–17]. All these measurements are sampled at 4 Hz before transmission to Earth, and are summarised in Fig. 1.

Given the complexity of the LISA instrument and operations [19], data artifacts are to be expected [20, 21]. Here, we will focus on the impact of missing data (that

* contact: ollie.burke@glasgow.ac.uk

† contact: martina.muratore@aei.mpg.de

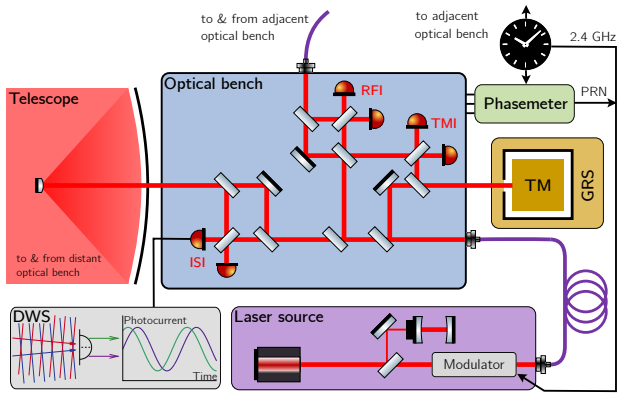


Figure 1. LISA’s measurement system, adapted from [18] with support from O. Hartwig. The Figure shows how each spacecraft uses three main interferometers (IFOs) to detect GWs. The Interspacecraft IFO captures the GW signal, buried within much louder noise, primarily from laser frequency fluctuations. The other two IFOs, the Reference and Test-mass IFOs, help remove noise due to spacecraft motion. Additionally, clock sidebands and pseudorandom noise codes correct for timing and laser noise, while the differential wavefront sensing measures beam angles to minimize angular jitter effects.

is, data gaps) potentially due to corrupted data that is removed having been rendered unsuitable for further analysis. The gap pattern considered in later sections of our study represent plausible scenarios that incorporate both scheduled short-term interruptions and unscheduled long-term data gaps.

Scheduled short gaps are expected when scientific measurements are temporarily disrupted by operational activities. For instance, these may occur during the periodic repositioning of the high-gain antenna to maintain Earth communication, or during adjustments made by the point-ahead angle mechanism (PAAM) to correctly align laser links. Although not strictly periodic, such gaps are anticipated to recur with near-regular frequency. Their duration may vary: antenna-related gaps could last a few hours and occur approximately every two weeks (medium duration gap), while PAAM-related gaps are expected to be shorter—typically under 100 s—occurring a few times per day.

In contrast, unscheduled long-duration interruptions are random and typically arise from unexpected anomalies. These may include events such as system failures requiring recovery procedures, similar to those encountered in LISA Pathfinder, or impacts from micrometeoroids. In the scenario adopted here, we assume several such events may occur annually, each lasting anywhere from one to three days.

Routine maintenance, as well as periods of partial or complete data loss are also expected. The nature and frequency of planned gaps will depend on the finalized

design and operational strategy of the spacecraft and instruments.

To the author’s knowledge, all existing studies on the impact of data gaps [20–26] have focused on signal recovery from post-processed gapped TDI observables, i.e., after the suppression of laser frequency noise, and none have examined the effects of missing phasemeter data on TDI itself. The goal of this work is to investigate how data gaps present in the scientific telemetry variables propagate through the initial noise-reduction pipeline and affect the construction of the final TDI observables. In particular, we aim to quantify the ‘augmentation’ of gaps: the increase in missing or invalid samples caused by the propagation of data gaps through intermediate processing steps. Understanding this compounding data loss is essential for ensuring that downstream stages of the LISA analysis pipeline, including parameter estimation and source population inference, are based on maximally preserved and well-characterised datasets. The global-fit for LISA aims to jointly recover both gravitational-wave sources and the underlying noise properties [27–30]. Accurate noise modelling is crucial since likelihood functions such as the Whittle likelihood [31] rely on assumptions of stationary, Gaussian, and circulant noise. Under these conditions, the Fourier transform diagonalizes the noise covariance, enabling efficient likelihood evaluation. However, real LISA data include effects like data gaps [32, 33] and non-circulant (Toeplitz-like) noise, making the Whittle likelihood only an approximation. Mischaracterising the noise can bias parameter estimation [21], so it is essential to carefully track and understand noise contributions from raw spacecraft measurements through to TDI variables and the global-fit likelihood. The results presented here aim to inform future strategies for telemetry management and gap mitigation.

This paper is organised as follows. In Sec. (II), we give a brief discussion of TDI (II A) and Lagrange-interpolation methods used to compute fractional delays (II B). Section (III) contains our main analytical results. Specifically, in subsection (III A) we discuss our analytical treatment of missing-data and gaps; and we compute the widening of the gaps in the intermediary variables in Sec.(III B), and in the TDI variables in Sec.(III C). Our analytical results are supported via appendices (A– C). Our results section is given in Sec.(IV). In Sec.(IV A) we numerically verify our analytical results in Fig. 2. To conclude, we then perform full LISA-like simulations in Sections (IV B) and (IV C) to understand (1) the duty cycle of the LISA mission and (2) the overall loss of data in a four-year LISA-like simulation due to TDI post processing on missing data. Finally, throughout our work, we will follow the conventions of PyTDI and LISA Instrument [34, 35].

II. TIME DELAY INTERFEROMETRY

To extract GW signals from LISA's interferometric data, it is essential to construct intermediate observables that suppress the most dominant noise sources (e.g. spacecraft motion (jitter) and laser frequency noise) to ultimately reduce the number of effective lasers from six to three. This is achieved through the derivation of intermediary variables, starting from the basic interferometric measurements collected onboard the spacecraft. Since there is already a plethora of material in the literature regarding TDI, we will report only a brief summary that is relevant for overall scope of this work. On this subject, we will not describe all instrumental noise sources which are both well-characterized and measurable. This includes, for example, the optical tilt-to-length coupling induced by spacecraft motion and clock noise (see [36], [37], and [38]). Residual noises refer to additional sources of noise that cannot be mitigated by the initial noise reduction pipeline. We can group these residual noises into two main types: a test-mass (TM) acceleration noise and a general optical metrology system (OMS) noise contribution associated with each individual inter-spacecraft link. We refer the reader to [39] for extensions involving multiple OMS noise components).

We will consider for our analysis only the three IFO measurements [40] (See Fig.1 for reference) per optical bench

- isi_{ij} — science IFO measuring the optical path between the laser hosted in the optical bench of the distance S/C j and the local laser i ,
- rfi_{ij} — reference IFOs interferes the local laser i and the laser in the adjacent optical bench j
- tmi_{ij} — test mass IFO interferes the laser in the adjacent optical bench j with the local laser i once this has bunched off the test mass.

Note that here, the indices ij refer to signals emitted by S/C j and received by S/C i . For instance, isi_{12} captures the signal transmitted from S/C 2 and S/C by craft 1.

A. Construction of Time-Delay Interferometry Variables

The first step in post-processing is the formation of the first class of intermediary ξ_{ij} variables, which are designed to suppress spacecraft jitter. These are computed as follows:

$$\xi_{ij} = \text{isi}_{ij} + \frac{1}{2}(\text{rfi}_{ij} - \text{tmi}_{ij}) + \frac{1}{2}\dot{\mathbf{D}}_{ij}(\text{rfi}_{ji} - \text{tmi}_{ji}), \quad (1)$$

where $\dot{\mathbf{D}}_{ij}$ is the Doppler-delay operator associated with the light travel time from S/C j to i .

For a generic time series $x(t)$, the Doppler-delay is applied as [41]

$$\begin{aligned} \dot{\mathbf{D}}_{ij}x(t) &= (1 - \dot{d}_{ij})\mathbf{D}_{ij}x(t) \\ &= (1 - \dot{d}_{ij})x(t - d_{ij}(t)) \end{aligned} \quad (2)$$

with $d_{ij}(t)$ denoting the (possibly time-varying) light travel time and $\dot{d}_{ij}$ the time-derivative of the light travel time between the two S/C j to i with arm-length L_{ij} with $d_{ij}(t) = L_{ij}(t)/c$. In our analysis, we are working in variables expressed in total frequency, rather than phase resulting in the use of Doppler-based delay operators [42].

Next, one constructs the second class of intermediary variables, η_{ij} , which serve to reduce the full system to three effective, independent laser noises:

$$\eta_{12} = \xi_{12} + \frac{1}{2}\dot{\mathbf{D}}_{12}(\text{rfi}_{21} - \text{rfi}_{23}), \quad (3a)$$

$$\eta_{13} = \xi_{13} + \frac{1}{2}(\text{rfi}_{12} - \text{rfi}_{13}). \quad (3b)$$

with four other variables η_{21}, η_{23} and η_{31}, η_{32} given by cyclic index permutations of the equations in (3).

To simplify the notation of sequential delays, we define chained delay operators [41]:

$$\dot{\mathbf{D}}_{i_1 i_2 \dots i_n} = \dot{\mathbf{D}}_{i_1 i_2} \dot{\mathbf{D}}_{i_2 i_3} \dots \dot{\mathbf{D}}_{i_{n-1} i_n}, \quad (4)$$

which recursively apply delays. For instance,

$$\dot{\mathbf{D}}_{ijk}x(t) = \dot{\mathbf{D}}_{ij}\dot{\mathbf{D}}_{jk}x(t) = [1 - \dot{d}_{ij}(t)][1 - \dot{d}_{jk}(t - d_{ij}(t))]x(t - d_{ij}(t) - L_{jk}(t - d_{ij}(t))),$$

demonstrating the non-commutative nature of Doppler-delay operators when arm lengths vary over time.

Using the η_{ij} variables, one can construct Michelson-

like TDI combinations that cancel laser frequency noise. First- and second-generation TDI observables, which incorporate increasingly longer sequences of delays to ac-

commodate time-dependent arm lengths, are given in factorized form:

$$\mathbf{X}_1 = (1 - \dot{\mathbf{D}}_{131})(\eta_{12} + \dot{\mathbf{D}}_{12}\eta_{21}) - (1 - \dot{\mathbf{D}}_{121})(\eta_{13} + \dot{\mathbf{D}}_{13}\eta_{31}) \quad (5)$$

$$\mathbf{X}_2 = (1 - \dot{\mathbf{D}}_{131} - \dot{\mathbf{D}}_{13121} + \dot{\mathbf{D}}_{1213131})(\eta_{12} + \dot{\mathbf{D}}_{12}\eta_{21}) - (1 - \dot{\mathbf{D}}_{121} - \dot{\mathbf{D}}_{12131} + \dot{\mathbf{D}}_{1312121})(\eta_{13} + \dot{\mathbf{D}}_{13}\eta_{31}). \quad (6)$$

The other TDI observables $\{\mathbf{Y}_1, \mathbf{Y}_2\}$ and $\{\mathbf{Z}_1, \mathbf{Z}_2\}$ are computed through cyclic permutation of the indices. The TDI observables form the backbone of LISA data analysis, enabling the extraction of GW signals through canceling the otherwise dominant laser noise. Many more TDI combinations can be built [43, 44], here we will only discuss the Michelson TDI because these are the current baseline used for LISA data analysis [45, 46].

B. Fractional Delays

Let $x(t)$ be a generic and continuous time-domain function. The discrete analogue of $x(t)$ is given by an evenly sampled sequence $x[n]$, with time-stamps $t_n = t_0 + n\Delta t$ with t_0 the initial starting time and $\Delta t = t_{n+1} - t_n$ the sampling interval. We define the observation time $T_{\text{obs}} = K\Delta t$ for $n \in [0, 1, \dots, K-1]$.

To ease notation in this section, we will only consider the usual delay operator $\mathbf{D}$ rather than the Doppler-delay $\dot{\mathbf{D}}$ since the term $\dot{d}_{ij}$ will vanish for our configuration as discussed later. Our results are easily extended for more general LISA-based orbit configurations, which we will leave for future work.

In general, due to the discrete nature of our time-series, the quantity $x(t - d_{ij}(t))$ will not exist as a discrete sample and must be approximated via interpolation schemes. The most common approach used throughout the literature is *Lagrange Interpolation* [47]. Lagrange interpolation has a maximally flat transfer function at $f = 0$, which implies that this interpolating method is very accurate for low frequencies where LISA will operate. A more efficient, and numerically more stable interpolant was recently developed in [48], which avoids (1) pathological behaviour at higher frequencies and (2) elevated number of terms (coefficients) in the interpolation scheme required to accurately infer the delay. The interpolating method found in [48] could easily be incorporated into this work, but will not change the results considerably.

Now, suppose the delay $d_{ij}(t) \in \mathbb{R}^+$. Then following [49], we can always make the split

$$d_{ij}(t) = d_{ij}^0(t) + d_{ij}^\epsilon(t), \quad (7)$$

for integer part $d_{ij}^0(t)$ and fractional part $d_{ij}^\epsilon(t)$. In our discrete notations, $d^0(t) \approx d_{ij}^0[n]$ with integer shift $N_{ij}^0 \in \mathbb{N}$ and fractional shift $N_{ij}^\epsilon[n] \in [0, 1)$ given by

$$N_{ij}^0[n] = \lfloor d_{ij}[n]/\Delta t \rfloor = \lfloor L_{ij}[n]/c\Delta t \rfloor \quad (8)$$

$$N_{ij}^\epsilon[n] = d_{ij}[n]/\Delta t - N_{ij}^0[n] \quad (9)$$

Notice that the integer part $N_{ij}^0[n]$ can be (1) time-varying due to the non-constant arm-lengths and (2) different (but constant) between individual links due to the arm-lengths being non-equal. Note that if we are to assume the individual links to be equal and non time-varying, resulting in a static orbital configuration, then $N_{ij}^0[n] = N$, i.e., a constant integer delay that is common across all arm-links.

In terms of continuous IFO measurements, we have

$$\begin{aligned} \mathbf{D}_{ij}\text{ifo}_{ij}(t) &= \text{ifo}_{ij}(t - d_{ij}(t)) \\ &= \text{ifo}_{ij}(t - d_{ij}^0(t) - d_{ij}^\epsilon(t)) \end{aligned} \quad (10)$$

which, in discrete notations, is given by

$$\begin{aligned} \mathbf{D}_{ij}x(t) &\approx (\mathbf{D}_{ij}x)[n] \\ &\approx \sum_{k=-p+1}^p \text{ifo}_{ij}[n - N_{ij}^0[n] + k] \ell_k(-N_{ij}^\epsilon[n]) \end{aligned} \quad (11)$$

with $p = (\mathcal{K} + 1)/2$, for $\mathcal{K}$ the order of the Lagrange interpolant and $\ell_r(x)$ the Lagrange basis functions defined by

$$\ell_r(x) = \prod_{j=-p+1}^p \frac{x - j}{r - j}, \quad j \neq r. \quad (12)$$

The expression of Eq. (11) is used to approximate the delay of a time-series at a particular (fractional) point. First and second-generation TDI variables defined by Eqs. (5) and (6) feature *nested* delays. Following [10], for simplicity, let $A_1, A_2, \dots, A_n$ represent potential S/C index pairings. Notice that

$$\begin{aligned} \mathbf{D}_{A_1 A_2 \dots A_n} x(t) &= \mathbf{D}_{A_1} \mathbf{D}_{A_2} \dots \mathbf{D}_{A_n} x(t) \\ &= x(t - d_{\text{total}}), \end{aligned}$$

with total time-delay

$$d_{\text{total}}(t) = d_1(t) + \mathbf{D}_{A_1} d_2(t) + \dots + \mathbf{D}_{A_1 \dots A_{n-1}} d_n(t),$$

that can be computed recursively in the following manner

$$d_{\text{total}}(t) = \sum_{k=1}^n z_k(t), \quad (13)$$

with

$$z_1(t) = d_1(t), \quad (14)$$

$$z_q(t) = d_q \left(t - \sum_{k=1}^{q-1} z_k(t) \right). \quad (15)$$

Equation (15) is then computed using Lagrange Interpolation. In the general orbital case where the LISA-arms are neither equal or constant, the delays are smooth functions with very low frequency content. For this reason, the recursive computation defined via Eq. (15) can be computed with very low interpolation orders that still results in highly accurate interpolated measurements [50].

In order to provide a simple analytic treatment that describes the impact of data gaps on TDI computations, we will consider a few simplifications to the orbital motion of LISA. We will assume that the arm-lengths of LISA are equal (and set to the average arm length of $\bar{L} = L_{ij} = 2.5 \cdot 10^9$ m) and constant ($\dot{L}_{ij} = 0$). We will neglect the effect of having three clocks with different time offsets and drifts, allowing us to discard the clock sideband measurements from the analysis. In the conclusion we will discuss how this choice could impact our results. Overall, we do not expect our results to change considerably for more complex orbits.

To conclude, for the remainder of the paper we will consider fractional delays acting on the time-series that can be approximated by

$$\begin{aligned} \mathbf{D}_{A_1} \dots \mathbf{D}_{A_m} x(t) &\approx (\mathbf{D}_{A_1} \dots \mathbf{D}_{A_m} x)[n] \\ &\approx \sum_{k=-p+1}^p x(n - N + k) \ell_k(-\epsilon), \end{aligned} \quad (16)$$

with

$$\begin{aligned} N &= \lfloor m \cdot \bar{d} \rfloor, \quad \text{Integer part} \\ \epsilon &= m \cdot \bar{d} - N, \quad \text{Fractional part} \end{aligned}$$

for m the number of delay operators and $\bar{d}$ the average light travel ($\bar{d} = 8.34$ s [51]) time defined between any two-links.

III. RESULT OF MISSING DATA ON TDI VARIABLES

A. Loss of Scientific Telemetry

In the context of GW parameter estimation and general data analysis, it is reasonable to assume that missing data can be reflected via imputation with zeros [21, 26]. This amounts to a multiplication of the

original time-series with a gating function, zero for $t \in T_{\text{gap}}$ and one otherwise. From a parameter estimation point of view, it has been shown that the “windowing” method returns identical statistical quantities (likelihoods, signal-to-noise ratios, information matrices etc.) to the masking methods in either the time or frequency domain (more details can be found in ref.[21]). After TDI has been performed, then the *global-fit* analyses should, in theory, be invariant to whether you apply windows or masks to your data stream, provided that the likelihood is consistent with the noise generation process.

For the *noise mitigation pipeline* this is not the case. We observed that treating missing-data in telemetry variables with zeros results in failure to cancel laser noise at the boundaries of the gaps. Imputing the scientific telemetry data with zero-values is particularly dangerous if there are non-trivial trends to the data stream since this would lead to spurious noise artifacts in the final TDI streams, potentially leading to pathological parameter inference during the global-fit scheme. A much cleaner and more elegant method is to assume complete ignorance of missing data – via imputation with unrepresentable and undefined floating points – NaNs¹.

For the following discussion, we will define

- **Valid Data:** The contiguous set of defined numerical data that are not corrupted by NaNs.
- **Missing Data:** Specific elements of data that are undefined and cannot be represented by floating point numbers. We will treat these numerically with NaNs.
- **Gap:** The contiguous set of missing data values between two sets of valid data (a missing packet of data is an element of a gap).

To simplify the discussion, we will assume that each of the IFOs have missing data at the same time-stamps. Similarly, for now, we will also assume that we only have one set of contiguous NaNs, implying a single gap in the data stream.

More formally, let $n = n_0 \in \{0, \dots, K-1\}$ be the first instance (and discrete index) of a missing data packet in a contiguous sequence of missing scientific telemetry data (IFO measurements) of length $N_{\text{NaNs}}^{\text{ifo}}$. This means that all of the IFO measurements will exhibit missing data for indices $n \in \mathcal{S}_{\text{NaNs}}^{\text{ifo}}$,

$$\mathcal{S}_{\text{NaNs}}^{\text{ifo}} = \{n_0, n_0 + 1, \dots, n_0 + N_{\text{NaNs}}^{\text{ifo}} - 1\}, \quad (17)$$

¹ Strictly speaking, missing data of this form are represented by “quiet” NaNs: missing-data that does not cause code to fail execution on detection of a NaN. As such, these quiet NaNs can be propagated through noise-mitigation pipelines with the idea of understanding how many extra quiet NaNs are detected downstream.

such that

$$\begin{aligned}
\text{ifo}_{ij}[n \in \mathcal{S}_{\text{NaNs}}^{\text{ifo}}] &= \text{isi}_{ij}[n \in \mathcal{S}_{\text{NaNs}}^{\text{ifo}}] \\
&= \text{tmi}_{ij}[n \in \mathcal{S}_{\text{NaNs}}^{\text{ifo}}] \\
&= \text{rfi}_{ij}[n \in \mathcal{S}_{\text{NaNs}}^{\text{ifo}}] \\
&= \text{NaNs}.
\end{aligned} \tag{18}$$

The string of equalities in Eq. (18) should be interpreted as a loss of communication between the instrument and ground-segment. This may not be the case in reality. For instance, we may find that the IFOs on board S/C 1 may exhibit significant glitchy behavior that results in total data loss, whereas the IFOs on S/C 2 and 3 operate at normal level. Until more realistic LISA-based simulations are conducted, we cannot derive a complete scenario of possible collections of missing data (gaps). For this reason, and since this work will form the basis for more detailed missing-data studies on noise mitigation pipelines (pre global-fit), we will assume that all the IFOs measurements are corrupted and lost at the same time-stamps. We will discuss how one could compute the extra missing data due to TDI calculations when considering more general scenarios in later stages of our work.

B. Gap Augmentation in the Intermediary Variables

The goal here is to understand how many extra missing data values are present in the intermediary variables η_{ij} when a fixed number of missing data exists in the interferometric measurements.

Recall from Eq. (18) that each of the IFO measurements exhibit the same missing data at the same time stamps. From the perspective of missing data values, this implies that the computed intermediary variables

η_{ij} will have missing data at the same time-stamps. In other words, defining the set $\mathcal{S}_{\text{NaNs}}^\eta$ as the ordered set of missing-data values in the intermediary variables, then from Eqs. (3) and (18), we must have for static instrument configurations that $\eta_{ij}[n \in \mathcal{S}_{\text{NaNs}}^\eta] = \text{NaNs}$ for all six S/C pairings. This means that we only need to focus on computing the gap augmentation in a single intermediary variable, since all other intermediary variables will host the same set of NaNs at precisely the same indices.

We can re-write Eq. (1) and Eq. (3) to focus on the individual components hampered by missing data. The goal here is to understand the gap augmentation of the following equation.

$$\hat{\eta}_{ij}(t) \approx \text{ifo}_{ij}(t) + \mathbf{D}_{ij} \text{ifo}_{ij}(t). \tag{19}$$

We name the standard η_{ij} variables as $\hat{\eta}_{ij}$ since we do not consider the specific numerical IFO measurements but the overall number of missing data (NaNs) in our calculations. This will simplify the expressions that feed into the TDI variables. In the discrete domain, we need to compute

$$\hat{\eta}_{ij}[n] \approx \text{ifo}_{ij}[n] + \sum_{k=-p+1}^p \text{ifo}_{ij}[n - N + k]. \tag{20}$$

Since we are only interested in the number of extra missing data, we can neglect $\ell_k(-\epsilon)$ since it will be unnecessary in our calculations.

Following Eq. (17) and Eq. (18), let $N_{\text{NaNs}}^{\text{ifo}}$ denote the number of NaNs in the telemetry variables. For both constant and equal arm-lengths, let $\bar{L}$ represent the length of a single LISA link, c the speed of light and Δt the sampling interval. The integer delay defined in Eq. (20) is then given by $N = \lfloor \bar{L}/(c\Delta t) \rfloor$. As derived in Appendix (A), the overall gap augmentation in the intermediary variables (Eq. (20)) is given by the piecewise function

$$N_{\text{NaNs}}^{\text{ifo} \rightarrow \eta} = f(\mathcal{K}, N_{\text{NaNs}}^{\text{ifo}}, N) = \begin{cases} \mathcal{K} + N_{\text{NaNs}}^{\text{ifo}} & \text{if } 1 < \mathcal{K} \leq 1 + 2N - 2N_{\text{NaNs}}^{\text{ifo}}, \\ \frac{\mathcal{K}+1}{2} + N & 1 + 2N - 2N_{\text{NaNs}}^{\text{ifo}} \leq \mathcal{K} \leq 2N - 1, \\ \mathcal{K} & \mathcal{K} \geq 2N - 1, \end{cases} \tag{21}$$

for $\mathcal{K}$ the order of the Lagrange interpolant for the η_{ij} variables. The overall number of missing-data values in the time-series is given by

$$N_{\text{NaNs}}^\eta = N_{\text{NaNs}}^{\text{ifo}} + N_{\text{NaNs}}^{\text{ifo} \rightarrow \eta}. \tag{22}$$

We remark in general that if one were to use the more numerically stable (and efficient) interpolant found in

Ref. [48], the result given in Eq. (21) will remain unchanged. The interpolating scheme in Ref. [48] requires far fewer coefficients (22) than the Lagrange interpolants used throughout this work (with default Lagrange interpolation order set to be 45). This would imply that the gap augmentation would be less, since we use less terms to approximate the delayed sample.

We remind the reader that the calculations leading to Eq. (21) have assumed a static and equal arm-length constellation. In the more general case of time varying arm-lengths, the value of the integer time-shift N will now depend on (1) a specific instance of time and (2) the S/C pair (ij). In other words, $N := N_{ij}^0[n]$ as reported in Eq. (11). The LISA arm lengths vary by approximately 3% for the entire course of the LISA mission [51]. As a result, depending on the chosen simulation start time and duration, the mean arm length can take values as large as $\bar{L}_{\max} = 2.54 \times 10^9$ m or as small as $\bar{L}_{\min} = 2.455 \times 10^9$ m. The integer part of the delay $N = \lfloor \bar{L}/c\Delta t \rfloor \in \{32, 33\}$ for $\bar{L} \in \{\bar{L}_{\min}, \bar{L}_{\max}\}$ respectively. In contrast, for our numerical simulations in our results section (IV A), for equal and constant arm-links with $\bar{L} = 2.5 \times 10^9$ m, one obtains $N = 33$ for the integer part of the delay. Making reference to Eq. (21), the integer time-shift primarily determines which of the three gap augmentation regimes that must be considered when predicting the overall widening of the gap. This means that our assertion that $\bar{L} = 2.5 \times 10^9$ m will manifest as an overestimation of the gap widening by a single data point per gap at most. This is in the very specific case that we have an order of interpolant $\mathcal{K}$ and original gap length $N_{\text{NaNs}}^{\text{ifo}}$ that satisfy the intermediary regime set in Eq. (21), the regime where the integer time-shift determines the size of the gap widening.

To conclude, we strongly believe that our results (and conclusions) on our scientific sections IV will not change if we were to consider more complex orbital configurations.

In the next section, we compute the overall missing data in the first and second generation TDI variables.

C. Gap augmentation in TDI Michelson variables

We will assume we have a single set of contiguous missing data in the telemetry variables, resulting in a single gap. We remind the reader since the telemetry variables all have the same missing data at the same time-stamps, we understand that each of the 6 intermediary variables will all have missing data at the same time-stamps. A result of this is that the missing data in the TDI variables $\{\mathbf{X}_1, \mathbf{X}_2\}$ will also be the same as $\{\mathbf{Y}_1, \mathbf{Y}_2\}$ and $\{\mathbf{Z}_1, \mathbf{Z}_2\}$. As a result, we will only provide the calculation for the TDI $\mathbf{X}$ variables, since the other variables for $\mathbf{Y}$ and $\mathbf{Z}$ follow trivially.

A computationally favorable scheme for computing both generation of TDI variables $\mathbf{X}_1$ and $\mathbf{X}_2$ defined in Eq. (5) and Eq. (6) is the following [52]. First we can define the quantities

$$a_1 = \eta_{12} + \mathbf{D}_{12}\eta_{21}, \quad (23a)$$

$$a_2 = \eta_{13} + \mathbf{D}_{13}\eta_{31}, \quad (23b)$$

that can be understood as single photon-path round-trip. Then defining second photon-path round-trip

$$r_1 = a_1 + \mathbf{D}_{121}a_2, \quad (24a)$$

$$r_2 = a_2 + \mathbf{D}_{131}a_1, \quad (24b)$$

it is possible to compute first generation TDI variables

$$\mathbf{X}_1 = r_2 - r_1. \quad (25)$$

Notice that to build $\mathbf{X}_1$ we require only 4 interpolation schemes rather than 6 required for the non-factorized expressions first generation TDI expressions [42, 47].

The second generation variables require a final set of variables

$$q_1 = (1 + \mathbf{D}_{12131})r_1, \quad (26a)$$

$$q_2 = (1 + \mathbf{D}_{13121})r_2, \quad (26b)$$

which can be combined to give the second generation TDI variables

$$\mathbf{X}_2 = q_1 - q_2, \quad (27)$$

which require only six interpolations, rather than 14 interpolations required to compute *unfactorized* TDI variables. This implies that the cost of computing the second generation factorized TDI variables is significantly less than the unfactorized quantities. We discuss in Appendix (B) how to compute the gap augmentation on factorized variables that make use of quantities such as

$$y(t) = x(t) + \mathbf{D}_{i_0 i_1 \dots i_n} x(t). \quad (28)$$

Notice that each of Eq. (23), Eq. (24) and Eq. (26) are of a similar form to Eq. (28) provided each have missing data at the same time-stamps (as assumed previously). For nested delays, we can use the formula (B3) defined in Appendix (B), which is a generalization of Eq. (21) for multiple delay operators. The goal then is to (repeatedly) compute the gap augmentation as Eq. (28) with the gap augmentation given by

$$N_{\text{NaNs}}^{x \rightarrow y} = f(\mathcal{K}, N_{\text{NaNs}}^x, \lfloor n \cdot \bar{L}/(c\Delta t) \rfloor),$$

for n the number of delay operators for each of Eq. (23), Eq. (24) and Eq. (26) and N_{NaNs}^x the number of missing-data present in the x variables is determined by Eq. (28).

We begin by computing the gap augmentation in each of the a_1 and a_2 single round-trip variables defined in Eq. (23a) and Eq. (23b). Since each IFO has missing data in the precise same time-stamps, this implies that the gap augmentation in a_1 should be identical to a_2 . The gap augmentation on both a_1 and a_2 is given by the following quantity

$$N^{\eta \rightarrow a_{(1,2)}} = f(\mathcal{K}, N_{\text{NaNs}}^\eta, \lfloor \bar{L}/(c\Delta t) \rfloor), \quad (29)$$

with overall number of missing data in both a_1 and a_2 , is given respectively by

$$N_{\text{NaNs}}^{a(1,2)} = N_{\text{NaNs}}^{\eta} + N^{\eta \rightarrow a(1,2)}, \quad (30)$$

with N_{NaNs}^{η} given by Eq.(22). In a similar manner, we can use Eq.(B3) to compute the gap augmentation for r_1 and r_2 . We thus have

$$N^{a(1,2) \rightarrow r(1,2)} = f(\mathcal{K}, N_{\text{NaNs}}^{a(1,2)}, \lfloor 2\bar{L}/(c\Delta t) \rfloor), \quad (31)$$

with total number of NaNs present in the r_1 and r_2 variables

$$N_{\text{NaNs}}^{r(1,2)} = N_{\text{NaNs}}^{a(1,2)} + N^{a(1,2) \rightarrow r(1,2)}. \quad (32)$$

For second generation variables, we have a final set of nested delays that require computing in Eq.(27). Given the overall number of NaNs present in the round-trip variables $N_{\text{NaNs}}^{r(1,2)}$, we compute

$$N^{r(1,2) \rightarrow q(1,2)} = f(\mathcal{K}, N_{\text{NaNs}}^{r(1,2)}, \lfloor 4\bar{L}/(c\Delta t) \rfloor) \quad (33)$$

$$N_{\text{NaNs}}^{q(1,2)} = N_{\text{NaNs}}^{r(1,2)} + N^{r(1,2) \rightarrow q(1,2)}, \quad (34)$$

where Eq.(33) determines the overall gap augmentation in the q_1 and q_2 variables. By construction, these have the same missing data placements as the second-generation TDI variables $\mathbf{X}_2$.

The quantities Eqs.(21), (29), (31) and finally Eq.(33) can be used to find out the total number of missing data in both first and second generation TDI variables. Combining each of these equations leads to the main result of our paper Eq.(35):

$$N_{\text{NaNs}}^{\mathbf{X}_1} = N_{\text{NaNs}}^{\text{tel}} + N_{\text{NaNs}}^{\text{tel} \rightarrow \eta} + N_{\text{NaNs}}^{\eta \rightarrow a(1,2)} + N_{\text{NaNs}}^{a(1,2) \rightarrow r(1,2)}, \quad (35a)$$

$$N_{\text{NaNs}}^{\mathbf{X}_2} = N_{\text{NaNs}}^{\mathbf{X}_1} + N_{\text{NaNs}}^{r(1,2) \rightarrow q(1,2)}. \quad (35b)$$

The two Eqs.(35a – 35b) determine the overall gap augmentation in both the first- and second-generation TDI variables.

IV. RESULTS

We begin by verifying our analytical results derived in Sec.(III C), specifically Eqs.(35), using publicly available LISA simulation codes. Once we show in Sec.(IV A) that our cheap-to-compute gap augmentation codes are faithful to LISA-based simulations, we will upscale our time-series and consider a full four year LISA-like data set with a realistic missing data and gap scenario in Sec.(IV C). We make use of our analytical results since with the current version of LISA instrument we cannot

run a four-year long simulation due to memory usage limitations. We will then compute the corrected duty-cycle when accounting for the gap augmentation due to TDI.

A. Verification

1. Simulation set-up

In this section, we will make use of `lisainstrument` to generate a small-scale LISA-based simulation for a static constellation with equal arm-lengths set to $\bar{L} = 2.5 \cdot 10^9$ m for each arm-link. We will consider a sampling rate $f_s = 4$ Hz and disable all noise sources except from laser, TM and OMS noise. The overall simulation will be around 8.3 hours in duration.

From the LISA simulation, we can extract the relevant scientific telemetry data. Specifically, we can obtain the measurements of the six interferometers for each craft. To construct segments of missing data, we will use a binary mask $w[n]$ with $w[n] = 1$ for $n\Delta t \notin T_{\text{gap}}$ and $w[n] = \text{NaN}$ for $n\Delta t \in T_{\text{gap}}$. Here T_{gap} is the time-interval of the missing data values. We enforce a gap in the telemetry variables by multiplying each of the ifo_{ij} measurements (isi_{ij} , tmi_{ij} and rfi_{ij}) by the masking function w . We do not apply a mask to the IFO sidebands and MPRs since we do not account for clock noise effects or evolving LISA arm-lengths.

These masked telemetry data are fed into `pyTDI` to compute the intermediary variables and resultant first and second generation TDI X variables. When computing the intermediary and TDI variables, we fix the order of interpolants to $\mathcal{K} = 45$. For a variable number of missing data (size of binary mask), we compare the result of our analytical calculation Eq.(35) with the numerical gap augmentation outputted using `PyTDI`.

In our analysis we do not correct for the group delay introduced when the phasemeter data are filtered and decimated from 80 MHz to 4 Hz. Compensating for this delay would require a time-shift of the telemetry, which in turn would increase the number of missing samples by an amount $\mathcal{A}$, the order of the Lagrange interpolant used for the shift [53]. These extra gaps would then propagate through (35) and (21). Since our focus here is on gap growth in the scientific data rather than the full noise-reduction pipeline, we ignore this effect; the additional widening is small and does not change our conclusions.

2. Discussion

Our result is given in Fig. 2. In all four panels we show that our analytical result of Eq.(35) is consistent

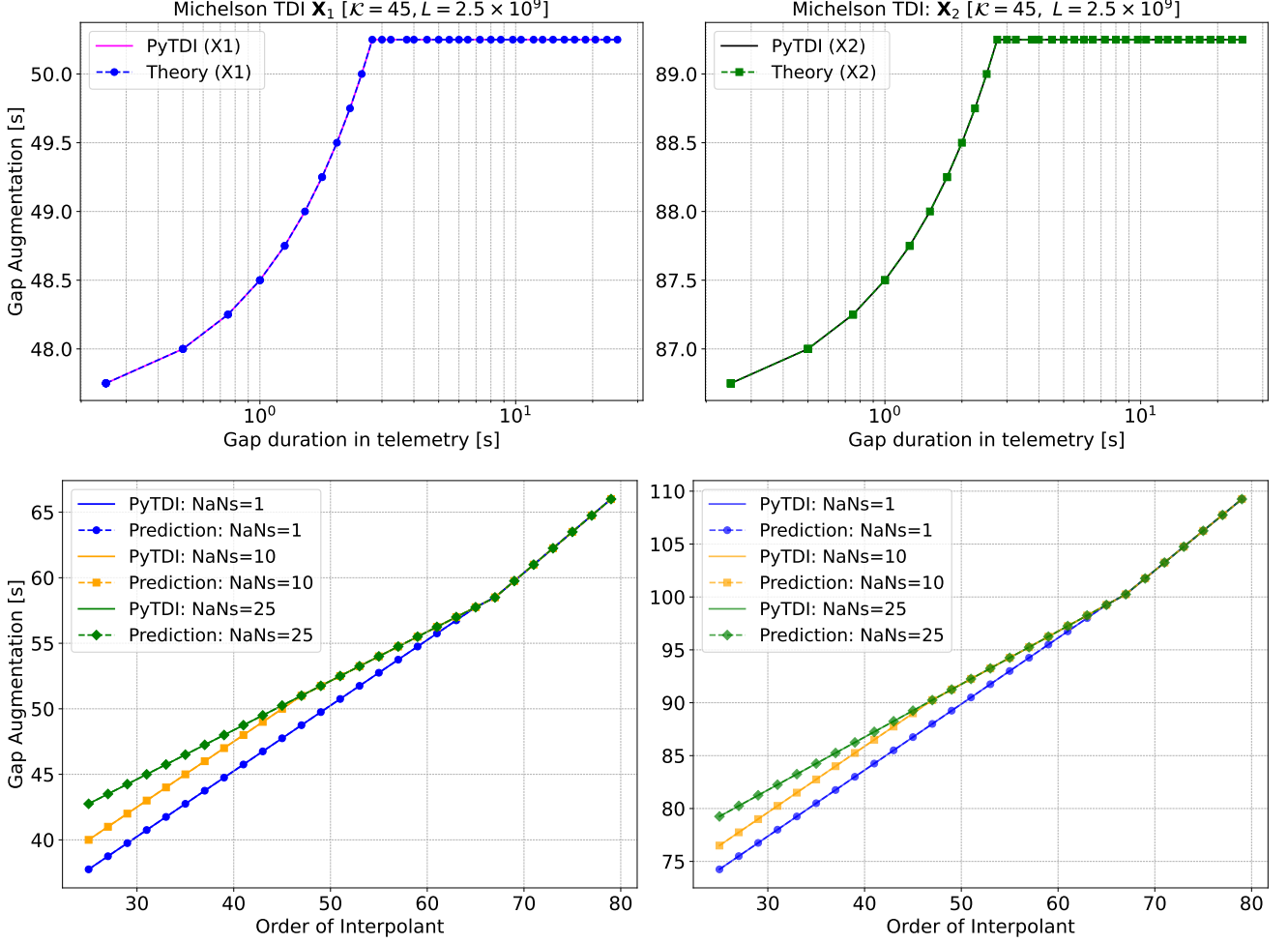


Figure 2. **(Top Left/Top right plot):** For first/second generation TDI variables (Eq.(25)) we compute the gap augmentation with $\mathcal{K} = 45$ for all Lagrange interpolants as a function of the gap duration. The purple/blue-dashed curves in the top left plot compare the numerical/theoretical gap augmentation in the first generation TDI variables using `lisainstrument` and `pyTDI` and Eq. (35a) respectively. The second plot is the same but for second generation TDI variables with black/green curves depicting the gap augmentation using `pyTDI` and Eq. (35b). **(Bottom Left/Bottom right plot):** Here we compare the Gap Augmentation using `lisainstrument` and `pyTDI` against our analytical result of Eq. (35a) (left plot) and of Eq. (35b) (right plot) as a function of the order of interpolant (set equal for all delay operators). As we can see, there exist three regimes where the gap augmentation changes slope with respect to the order of interpolant (see Eq. (21) for similar behaviour with justification outlined in Appendix B). As one can see, our analytical result (dashed blue, yellow and green dots) matches the numerical simulations (solid blue, yellow and green curves) for both plots for various sets of contiguous nans $N_{NaNs}^{ifo} = \{1, 10, 25\}$. A `jupyter` notebook with code to reproduce these plots is available [here](#).

with the numerical gap augmentation given using both `lisainstrument` and `pyTDI`. In the upper plots of the two panels in Fig. 2, we report the gap augmentation with respect to the number of NaNs (or equivalently to the duration of the gap) when a fixed interpolation order of 45 has been chosen. Moreover, on the bottom of the two panels we also analyse three scenarios with 1, 10, and 25 NaNs and we study the gap augmentation with respect to the order of interpolation.

On the top two plots, for both first- and second-

generation TDI variables ($\mathbf{X}_1$ and $\mathbf{X}_2$), for gaps larger than 2.5 s (or equivalently for number of NaNs > 10), the augmentation becomes constant and equal to

- 50.25 s for first generation variables,
- 89.25 s for second generation variables.

This is interesting since it shows that the gap augmentation is not strongly dependent on the gap duration; e.g., for gap durations ≥ 2.5 s, the gap augmentation appears to be constant.

Frequent and short gaps may be the most concerning. A loss of a single packet of data (~ 0.25 s) can lead to an additional loss of ~ 86.75 s in TDI $\mathbf{X}_2$. This is particularly relevant for LISA if very short but highly frequent data gaps occur (e.g., antenna-related gaps last about a few hours while PAAM-related gaps are typically of 100 s), as the cumulative excess loss from TDI could become significant. **By contrast, less frequent and longer gaps are less affected by TDI.** While TDI $\mathbf{X}_2$ can extend a gap by ~ 90 s, this is negligible compared to scheduled long gaps (e.g., estimation of TTL coupling coefficients), which can last up to ~ 40 hours. Although any additional loss of data is undesirable, its relative impact in this case is minimal. **In summary**, when comparing losses from TDI operations to those caused by the gaps themselves, short and frequent gaps pose the greatest risk. For example, short gaps are expected to occur frequently in the planned schedule (1095 per year, ~ 100 s each), making the TDI-induced loss comparable to the intrinsic loss. Conversely, for long gaps, which are rare (6 per year), the additional TDI loss is negligible compared to their planned duration.

We will now conclude this section with a small discussion focused on more generic orbital motions of the craft. As discussed previously, the LISA arm lengths vary by at most $\sim 3\%$, resulting in an integer time-shift (see Eq. (21)) $N \in \{32, 33\}$. The situation becomes a touch more interesting when considering multiple delays, with overall delay expressed by

$$\left\lfloor \frac{m\bar{L}}{c\Delta t} \right\rfloor. \quad (36)$$

For multiple delays that determine the pre-factor (n) in Eq. (36), the choice of $\bar{L} \in [\bar{L}_{\min}, \bar{L}_{\max}]$ will govern smaller/larger values of the integer time-shift. Where here $\bar{L}_{\min}$ and $\bar{L}_{\max}$ can be estimated from the pre-computed orbits given in Ref [51]. The value of this integer time-shift determines the regime of gap augmentation located in Eq. (B3). Note that the only regime where the integer time-shift features directly in the gap augmentation is the secondary regime in Eq. (B3). For second generation TDI variables, we can compute $\lfloor 4\bar{L}/(c\Delta t) \rfloor \in [131, 135]$ for $\bar{L} \in \{\bar{L}_{\min}, \bar{L}_{\max}\}$, to compare our case, when assuming $\bar{L} = 2.5 \times 10^9$ m, this results in $\lfloor 4\bar{L}/(c\Delta t) \rfloor = 133$. As one can see, the gap augmentation could either increase/decrease (by only a few points) only if our order of interpolant, and number of NaNs present in any of the steps previous to the TDI calculations, force us to fall within the intermediary regime (see for example r, a, η in Eqs.(21), (29) and (31)).

We have indeed checked our analytical formula for more generic orbits of LISA using `lisainstrument` and

`pyTDI` with $\bar{L} \in [\bar{L}_{\min}, \bar{L}_{\max}]$ and our observations were expected. For $\mathbf{X}_1$ and $\mathbf{X}_2$, we found that we underestimated the overall gap augmentation by around one and three points respectively, if $\bar{L} = \bar{L}_{\max}$ or $\bar{L} = \bar{L}_{\min}$ for a short ~ 8.3 hour simulation. For longer LISA-based simulations, the arm-lengths are slowly oscillating around a mean-value so the gap augmentation can decrease as well as increase. We believe that our average value is a solid baseline to make scientific conclusions out of. As a result, we strongly believe that our conclusions would not change significantly if we were to consider more general orbits of the LISA constellation.

B. LISA-based Gap Simulations

It is expected that LISA should operate normally for four years with a quoted $\sim 82\%$ of usable data, the so-called *duty cycle* of the mission [8]. Generating a full four-year long LISA simulation sampled at 4 Hz using both `lisainstrument` and `pyTDI` is not only extremely expensive but prohibitively memory intensive. Instead, using the results of the previous section and the details outlined in Appendix (C), one can compute the overall extension of missing data in TDI variables relatively cheaply. Our results can inform performance and operation personnel on the loss of data in case gaps are present in the raw phasemeter data and TDI-augmented.

Tab.(I), inspired by Ref.[20] and extended to include additional gap scenarios arising from the design of LISA, outlines a potential situation in which the spacecraft is affected by both *planned* and *unplanned* data gaps. Here, we define *planned* gaps as those due to routine instrumental maintenance, whereas *unplanned* gaps correspond to unexpected instrumental malfunctions or complete communication losses caused by environmental effects, leading to a total loss of usable data (i.e., not attributable to routine maintenance).

Such malfunctions can either produce excessive noise bursts that are subsequently removed from the data (glitch subtraction) or cause communication outages leading to the loss of interferometric variables required for TDI computations. These artifacts may result in corrupted data that cannot be used in parameter inference pipelines. While such data segments can be entirely removed (masked), this approach enables the application of simpler and more cost-effective search and parameter estimation techniques, as demonstrated in [20, 21, 23].

We sample the *planned* data gaps at regular intervals following the Rate [yr^{-1}] given in Table (I), with a $\sim 10\%$ random jitter applied to (the start time) the gap duration of each subtype of gap. For the *unplanned* data gaps, the gap duration (start times) are sampled from an exponential distribution which is derived again

Gap Type	Subtype	Rate [yr ⁻¹]	Duration [hr]	(masked) Segments	Total Hours	Duty Cycle [%]
Planned	Long	6	40	4339	1363.38	96.112
	Medium	26	3.3			
	Short	1095	0.028			
Unplanned	Platform	3	60	126	4156.82	88.145
	Payload	4	66			
	Environmental	30	24			
Planned + Unplanned	—	—	—	3944	5367.04	84.257

Table I. The first column represents the class of gaps, be them Planned, Unplanned or a mixture of both. The second column represents the subtype of gap with third and fourth columns giving expected rates/year and durations [hours]. The fifth, sixth and seventh columns represent the median number of gap segments, total hours of missing data and overall duty cycle for ~ 2000 seeded masking simulations sampled at $\Delta t = 10$ s with assumed four-year mission duration. The median total duty cycle reflects the proportion of the year uncontaminated by gaps. The distributions of the duty cycles are given in Fig. 3. We note that the total number of gated segments in the planned+unplanned configuration is smaller than in the planned-only case, since some of the more frequent but shorter planned data gaps are absorbed into the longer and less frequent unplanned gaps.

accordingly to the Rate [yr⁻¹] given in Table (I). The result is a binary mask with integer valued 1 representing valid data and NaNs representing missing data.

In the following, we will assume that the instrument loses data in all IFO measurements at precisely the same time-stamps. We will construct three masks, one for either planned, unplanned and with both planned and unplanned with rate parameters and durations given by the second and third column of Tab. (I).

Based on approximately 2000 seeded realizations of the four-year masking function, sampled at $\Delta t = 10$ s with the parameters specified in Tab.(I); Tab.(I) summarizes the median number of masked segments, total hours lost, and duty cycles. The distributions on the telemetry duty cycles is given in Fig.(3).

What we learn from Tab. (I) and Fig. 3 is that:

- Unplanned data gaps account for the majority loss of data for the instrument, accounting for median $\sim 11.855\%$ loss, where this percentage is computed considering 88.145% of duty cycle,
- Planned data gaps only account for a median $\sim 3.888\%$, where this percentage is again computed considering 96.112% of duty cycle,
- Therefore, the overall (median) duty cycle is $100\% - (11.855 + 3.888)\% = 84.257\%$; which is bigger than the quoted 82% nominal duty cycle [8, 20].

This latter point means that we retain approximately $2.257\% \approx 35.61$ days of worth of scientific data compared to the previously estimated (nominal) 82% duty cycle [8, 20]. **The nominal 82% duty cycle only accounts for the individual gaps affecting the LISA**

data rather than their combination², which is outlined in Tab. (I). This would result in an **increase of ~ 35.61 days to garner more science from GW observations.**

C. Full scale gaps in second generation TDI Michelson variables

We can go further and compute the extra loss of data due to TDI data streams with missing data. We begin by simulating 600 seeded data gaps with rate parameters defined in Tab. (I), at a sample rate of 4 Hz. To compute the overall number of missing data on telemetry variables, with arbitrary numbers of gaps and of arbitrary duration, we refer the reader to Appendix (C). Specifically, we will use the equations that lead up to Eq. (C9) to compute the widening of the individual gaps for second generation Michelson TDI variables. Our median results are tabulated in Tab. (II).

1. Unplanned Gaps

In this paragraph we will see that although the unplanned data gaps are comparatively much larger than the planned data gaps (see Table (I)), the impact of the extra missing data governed through TDI is far less severe than the planned gaps. Results are summarized in

² Note that short duration, high rate gaps may be contained in long-duration, low rate gaps as extensively demonstrated in Appendix (A) and Appendix (B).

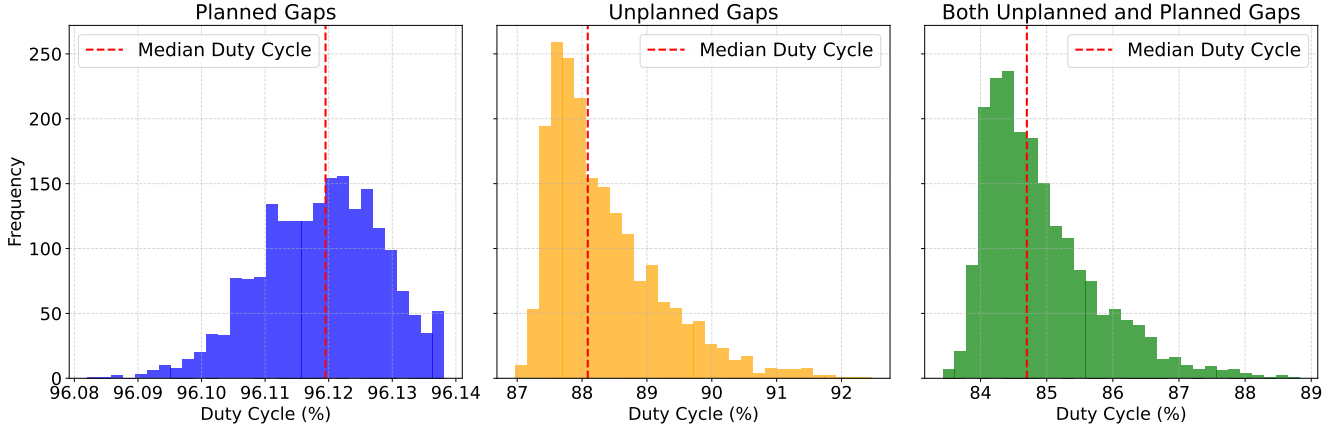


Figure 3. The left/center/right plots show distributions on the resultant duty cycles on the telemetry data when planned/unplanned/both planned and unplanned gap strategies are used via the rates in Tab. (I).

Table (II). The first column reports the type of gap considered, the second column gives the duty cycle of the telemetry data, the third column shows the duty cycle after applying TDI — where some data are lost due to gap augmentation — and the last column presents the difference between the two duty cycles, i.e., the total data loss over the 4-year observation period.

For unplanned data gaps we lose out on ~ 3 hours worth of data after applying TDI (see Tab. (II)), a small fraction when compared to the overall loss of data due to the unplanned data gaps. Since the order of interpolant has been fixed to $\mathcal{K} = 45$, the gap widening becomes constant for gaps longer than 2.5 s. This has been previously depicted in Fig. 2, which shows that the gap augmentation becomes constant for gap durations longer than seconds.

Although the gaps are large comparative to the planned data gaps in duration, the gap widening is constant and small, resulting in an augmentation (after TDI has been applied) of ~ 90 s per gap.

Given the family of unplanned gaps tabulated in Tab. (I), when not accounting for the merging of the gaps for the unplanned data set, we would expect approximately $4 \cdot (30 + 4 + 3) \cdot 89.25 \text{ s} \sim 0.15$ days lost of data where we have considered the total Rates, the gap augmentation of 89.25 s for TDI $\mathbf{X}_2$ and 4 years of observation. This is a reasonable (but overestimated) approximation when compared to Tab.(II) that shows the true gap augmentation in the TDI variables that accounts for potential merging of gaps.

2. Planned Gaps

For the planned data gaps we see that after applying TDI the overall loss of data is much larger than the

unplanned ones, approximately ~ 4.4852 days worth, which is approximately ~ 35 times more data lost than the unplanned data gaps. This is primarily due to the high-rate and short-duration data gaps with rate 1095/yr. Indeed, from a back-of-the-envelope calculation we see that there should be $4 \cdot (6 + 26 + 1095) \cdot 89.25 \text{ s} \approx 4.66$ days worth of lost data. We see from Tab. (II) that our estimate is reasonable, with the true result for the total lost LISA data for the planned scheme being ≈ 4.48 days, where here we have accounted for gap merging. Finally, the combination of both planned and unplanned data gaps results in a loss of duty cycle of approximately 0.28%, which is ~ 4 days worth of data.

To conclude, what we learn here is that gaps of short duration but high rates of occurrence will result in the greatest amount of extra lost data when ran through TDI-based noise-mitigation pipelines. Our results here support the literature that short and frequent gaps may result in a greater impact on the science for LISA than longer duration, but lesser frequent gaps.

V. CONCLUSIONS AND FUTURE WORK

The primary focus of this analysis is to understand the impact that missing raw phasemeter (scientific telemetry) data has on the resultant TDI observations. By treating missing instrumental telemetry data (and sets of missing data – gaps) in the form of unquantifiable floating points (NaNs), we have shown that ignorantly masking the telemetric data will result in a gap augmentation in the resultant first/second generation TDI streams for X, Y, Z . Since the delay operators are computed using FIR (Finite Impulse Response filter) filters,

Gap Type	Duty Cycle (Telemetry)[%]	Duty Cycle (TDI) [%]	Data Lost [days]
Planned	96.12	95.81	4.4852
Unplanned	88.02	88.01	0.1291
Planned + Unplanned	84.65	84.37	4.0725

Table II. The first column represents the Gap type with gap parameters defined in Tab.(I). The second column is the median duty cycle in the masking function used to mask the telemetry variables. The third column is the overall duty cycle in the TDI variables after the gap augmentation has been taken into account. The final column is the potential loss of science data due to propagating missing-data from the scientific telemetry to the final TDI variables. In this simulation, we use 600 seeded telemetry-based masking functions sampled at 4 Hz with assuming a mission duration of 4 years.

we found that the gap augmentation linearly depends on the order of the kernel (and thus the order of the Lagrange interpolant used in this work). Analytical results were obtained (see Eq. (35)) that can cheaply compute the gap augmentation in the intermediary and first/second generation TDI Michelson variables ($\mathbf{X}_1$ and $\mathbf{X}_2$) for any family of gaps with arbitrary durations and rates. For first and second generation TDI variables, with a fixed order of 45 Lagrange interpolant and gap duration ≥ 2.5 s, we observe losses of data of duration ~ 90 s and ~ 50 s respectively. The loss of data is greatest for second generation TDI variables, rather than first generation TDI variables, and this is due to the overall number of fractional delay computations (and thus interpolants) applied to the data (with 6 in total applied to second generation rather than 4 for first generation). In Fig. 2, we found that the gap augmentation for first and second generation variables is largely insensitive to the gap duration for data gaps of duration longer than 2.5 s. As demonstrated in Fig. 2, we found that a single packet of lost telemetry data results in a loss of TDI-based data lasting approximately ~ 86.25 s. This highlights the severity of short but frequent gaps in the raw phasemeter data, which can lead to substantial data loss once propagated through TDI.

With analytical results in hand that can cheaply and accurately predict the gap augmentation in TDI variables, we then investigated a LISA-based simulation study on data gaps assuming a realistic gap scenario given by Tab. (I). *We demonstrated that the median effective duty cycle is 84.6% when accounting for both planned and unplanned data gaps, which is approximately $\sim 2.6\%$ higher than what has been quoted in the literature [1].* This is due to the realistic case of overlapping gaps in both the planned and unplanned configurations. Averaging over multiple realisations of four year long binary masks sampled at 4 Hz, we used our cheap-to-evaluate analytical results to quote median gap augmentation results in the second generation TDI variables. Our results were given in Tab. (II). We concluded that for planned gaps short duration but high rate data gaps lead to the greatest loss of data, resulting in an effective duty cycle loss of $\sim 0.3\%$, which is approximately ~ 4 days worth of lost data.

This loss of data due to TDI is not as severe for the unplanned data gaps (gaps with smaller rates but greater durations), with an overall loss of $\sim 0.01\%$, i.e., 3 hours worth of lost data.

High rate and short duration gaps that may appear harmless upstream in the noise mitigation pipeline may have drastic consequences downstream during the global fit analyses. Not only is the loss of scientific data never warranted, but these gaps could also have serious consequences on the search and parameter estimation schemes. Parameter estimation schemes rely on a solid understanding of the statistical properties of the noise process. The presence of many short-duration and high-rate gaps may start to impact assumptions on the process – in particular whether the between gap segments are independent, dependent. Spectral leakage, if not fully accounted for in the noise covariance matrix could also significantly increase the uncertainty in the parameter posterior [21, 24, 32, 33]. No matter, what domain, be it time, frequency or time-frequency [54, 55], all domains of analyses would be negatively impacted by these short duration gaps.

A clear take home here is that shorter and more frequent gaps will return a greater loss of data than longer but less frequent gaps (with the same duty cycle as the former). An outcome of this research is that, if possible, one should maintain the instrument in such a way to reduce the number of frequent gaps that could occur.

There are a number of future projects that could stem from the details contained in this work. For instance, we study here only the current baseline Michelson TDI X. However, other TDI combinations, such as the Sagnac type [43, 56, 57], might behave differently because they involve shorter delays and therefore gaps are less spread. Similarly, techniques like TDI-Infinity could be better suited to avoid losses due to gaps [58]. Moreover, the top right panel of Figure (2) demonstrates for kernel orders $\mathcal{K} = 45$ that if a single packet of missing data is lost (of duration ~ 0.25 s), then one could lose an significant amount of data in comparison (say, ~ 86.25 s worth). Here we provide a worse case scenario where a single missing data point results in a major loss of data. However, one could use more sophisticated interpolation techniques that are able to

account for *unevenly* sampled data points. Smaller sets of contiguous missing-data could be interpolated, or potentially imputed using probabilistic techniques. Interpolating the missing data will force the time-series to be evenly sampled, facilitating the use of spectral methods for downstream signal extraction analyses.

However, it is of paramount importance that the key idea of TDI is maintained: the overall subtraction and mitigation of laser noise that enables gravitational-wave inference. In practice, this means that any interpolation (or imputation) method is used to recover missing samples must be accurate enough to preserve the laser-noise cancellation. For example, if a single data point is missing and we replace it using interpolation, the reconstructed value must be consistent with what would have been measured so that the subsequent TDI combinations still achieve near-complete laser-noise removal. At the same time, interpolation is performed not only over noise but also over the underlying gravitational-wave signal. If the interpolation is inaccurate, it can distort the signal locally and introduce systematic biases in the recovered waveforms and astrophysical parameters. Therefore, interpolation schemes must simultaneously ensure robust laser-noise cancellation and safeguard the fidelity of gravitational-wave inference.

In this work, we have assumed that all telemetry losses occur at the same time-stamps across the interferometers, i.e., all observations are lost simultaneously. In reality, however, data gaps may occur at different times in different interferometer measurements. A natural extension of this study would be to explore the impact of such asynchronous data losses on TDI computations. In this case, the action of delay operators on the intermediary variables would cause each individual gap to widen differently in each interferometer stream. As these propagate downstream, the resulting TDI observables would exhibit multiple gaps, distributed across the X, Y, Z combinations at different time intervals. We expect that this scenario would lead to a greater effective loss of data in the TDI outputs, since the gap segments in different channels are no longer fully overlapping and may become mutually exclusive. From an implementation perspective, this effect could be represented as three distinct binary masks—one for each TDI variable—each containing gap segments at different time positions and with varying durations. This could be easily explored using the LISA simulation tools `lisainstrument` and `pyTDI`. We purposefully chose not to explore this further without a clearer understanding of the LISA instrument and what could result in such losses of information.

Lastly, the impact of clock noise on TDI processing – which is closely tied to the specifics of the data-processing pipeline – and therefore on gaps propagation has not been taken into account here. In the case of long-term clock drifts, the relative frequency offsets be-

tween spacecraft clocks can accumulate to significant values over the mission lifetime. For example, a relative offset of order 10^{-6} leads to a timing error of $\mathcal{O}(100)$ s over four years ($\sim 10^8$ s). While such drifts appear as common-mode terms in the applied delays and time-stamp offsets and therefore do not strictly need to be corrected prior to TDI, they would result in delays well outside the nominal arm-length scale $\bar{L}$ adopted throughout this work. It is thus preferable to correct for these large drifts, ensuring that the delays remain within a physically meaningful and constrained range. The in-band impact of clock correction depends on the chosen implementation strategy. Two main approaches are considered in the literature: (i) applying the correction jointly with the primary delays and Doppler shifts when constructing laser-noise-reducing TDI combinations [42], or (ii) applying it in a subsequent step after laser-noise reduction [38]. In either case, the required clock measurements are typically reconstructed through sensor fusion of sidebands, carrier tones, and possibly additional references such as pseudo-random noise codes, ground-based signals, or TDI-Ranging [59, 60]. The propagation of telemetry gaps through this reconstruction is non-trivial. For instance, in a simple sensor-fusion scheme, the sideband frequency fluctuations are integrated to recover the clock signal; if a data gap interrupts this process this force the integration to restart.

This work forms the foundations of missing-data on space-based noise-mitigation pipelines, and so we leave these concepts as future work.

VI. DATA RELEASE

The authors feel strongly about providing open-source software that others can use to build upon this work. The work here was not possible without the use of open sources codes – in particular `lisainstrument`, `pyTDI` and `lisaglitch` [34, 35, 61]. We release our codes in a Github repository found at [this link](#).

VII. ACKNOWLEDGEMENTS

O.Burke is deeply thankful for Eleanora Castelli for both the wonderful and insightful conversations regarding LISA simulation and for giving O. Burke access to her gap masking code. O. Burke also appreciates the many conversations he had with L. Speri and C. Chapman-Bird (who read and gave excellent comments on an early draft), particularly on gaining perspective on this work. M.Muratore is very grateful for the support given by Olaf Hartwig towards providing a better understanding of the results obtained in this manuscript and to P. Wolf and A. Hees for providing insight on the

effect of having gaps when clock noise is taken into account. M.Muratore also acknowledges previous fruitful discussions with A. Hees on gap propagation in TDI during the preparation of the document for the LISA Mission Adoption Review. Last, but certainly not least, M.Muratore acknowledge Wolfrang Kastau for his hint on using NaNs to propagate and keep track of missing data. O.Burke and M.Muratore are deeply thankful for many insightful discussions with the CU L01 group. In particular, O. Burke and M. Muratore are very grateful for the support given by Jean Baptiste-Bayle and Martin Staab. M.Muratore. gratefully acknowledges the support of the German Space Agency, DLR. The work is supported by the Federal Ministry for Economic Affairs and Climate Action based on a resolution of the German Bundestag (Project Ref. No. FKZ 50 OQ 2301). O. Burke and G. Woan acknowledge financial support from the Grant UKRI972 awarded via the UK Space Agency. This work has made use of `pyTDI`, `lisainstrument` and `lisaglitch`.

Appendix A: Derivation of Gap Augmentation for Intermediary Variables

Preliminaries

In our prescription, missing data will be represented via symbolic NaNs, where any binary operation (addition, multiplication, division, subtraction) involving NaNs will return NaNs. Let $n = n_0$ be the first instance (and discrete index) of a missing data packet in a contiguous sequence of missing data of length $N_{\text{NaNs}}^{\text{ifo}}$.

This means that all of the IFO measurements will exhibit missing data for indices $n \in \mathcal{S}_{\text{NaNs}}^{\text{ifo}}$ with ordered set

$$\mathcal{S}_{\text{NaNs}}^{\text{ifo}} = \{n_0, n_0 + 1, \dots, n_0 + N_{\text{NaNs}}^{\text{ifo}} - 1\} \quad (\text{A1})$$

with

$$\begin{aligned} \text{ifo}_{ij}[n \in \mathcal{S}_{\text{NaNs}}^{\text{ifo}}] &= \text{isi}_{ij}[n \in \mathcal{S}_{\text{NaNs}}^{\text{ifo}}] \\ &= \text{tmi}_{ij}[n \in \mathcal{S}_{\text{NaNs}}^{\text{ifo}}] \\ &= \text{rfi}_{ij}[n \in \mathcal{S}_{\text{NaNs}}^{\text{ifo}}] \\ &= \text{NaNs}. \end{aligned} \quad (\text{A2})$$

We will define

$$n_{\text{first}}^{\text{ifo}} = n_0 \quad (\text{A3})$$

$$n_{\text{last}}^{\text{ifo}} = n_0 + N_{\text{NaNs}}^{\text{ifo}} - 1 \quad (\text{A4})$$

as the first and last indices of the gapped segment in the telemetry variables. The goal now is to count the missing-data extension (gap augmentation) of the intermediary variables defined in Eq. (20). We are not interested, quantitatively speaking, on the specific values

of valid data in the intermediary variables quantitatively speaking. For this reason, we can neglect terms such as $\ell_k(-\epsilon)$ in Eq. (11). Given that

$$(\mathbf{D}_{ij}\text{ifo}_{ij})[n] \approx \sum_{k=-p+1}^p \text{ifo}_{ij}[n - N + k], \quad (\text{A5})$$

then the first and final indices (ordered left to right) of the NaN placement in Eq. (A5) are given respectively by

$$n_{\text{first}}^{\mathbf{D}_{ij}(\text{ifo})} = n_0 + N - p \quad (\text{A6a})$$

$$n_{\text{last}}^{\mathbf{D}_{ij}(\text{ifo})} = n_0 + (N_{\text{NaNs}}^{\text{ifo}} - 1) + N - (1 - p) \quad (\text{A6b})$$

We can thus write the set of NaN index placements in Eq. (A5)

$$\mathcal{S}_{\text{NaNs}}^{\mathbf{D}(\text{ifo})} = \{n_{\text{first}}^{\mathbf{D}_{ij}(\text{ifo})}, n_{\text{first}}^{\mathbf{D}_{ij}(\text{ifo})} + 1, \dots, n_{\text{last}}^{\mathbf{D}_{ij}(\text{ifo})}\}. \quad (\text{A7})$$

Notice that the number of contiguous NaNs between the first and last indices (via an inclusive difference) of the equation above is

$$(n_{\text{last}}^{\mathbf{D}_{ij}(\text{ifo})} - n_{\text{first}}^{\mathbf{D}_{ij}(\text{ifo})} + 1) = \mathcal{K} + N_{\text{NaNs}}^{\text{ifo}}. \quad (\text{A8})$$

The goal now is, given a number of contiguous NaNs in the set $\mathcal{S}_{\text{NaNs}}^{\text{ifo}}$ defined in Eq. (A1), to compute the gap augmentation in the expression

$$\hat{\eta}_{ij}[n] = \text{ifo}_{ij}[n] + \mathbf{D}_{ij}\text{ifo}_{ij}[n]. \quad (\text{A9})$$

The overall extension of the telemetry missing data (gap augmentation) depends on the overall missing data placement in both the sets $\mathcal{S}_{\text{NaNs}}^{\text{ifo}}$. As we will see in the next few subsections, there are three regimes that must be considered that depend on the order of interpolant $\mathcal{K}$, integer time shift N and number of contiguous missing data in telemetry $N_{\text{NaNs}}^{\text{ifo}}$.

Regime 1: No Merging

Let us suppose that the NaN placements in the telemetry variables do not have the same time-stamps as the delayed telemetry variables defined in (A5). Mathematically, this can be written using the set notation defined in the previous section

$$\mathcal{S}_{\text{NaNs}}^{\text{ifo}} \cap \mathcal{S}_{\text{NaNs}}^{\mathbf{D}_{ij}(\text{ifo})} = \emptyset. \quad (\text{A10a})$$

In other words, the two sets defined by Eq. (A1) and (A7) are *mutually exclusive* and have no common elements. This means that the extra missing data induced via the delay operator is given by the size of the set

(A7). As shown in (A8), the gap augmentation $N_{\text{NaNs}}^{\text{tel} \rightarrow \eta}$ is given by

$$N_{\text{NaNs}}^{\text{tel} \rightarrow \eta} = \mathcal{K} + N_{\text{NaNs}}^{\text{ifo}}. \quad (\text{A11})$$

With total number of missing data in the intermediary $\hat{\eta}_{ij}[n]$ variables given by

$$\begin{aligned} N_{\text{NaNs}}^{\eta} &= \underbrace{N_{\text{NaNs}}^{\text{ifo}}}_{\text{Telemetry}} + \underbrace{\mathcal{K} + N_{\text{NaNs}}^{\text{ifo}}}_{\text{Gap Augmentation}} \\ &= \mathcal{K} + 2N_{\text{NaNs}}^{\text{ifo}}. \end{aligned} \quad (\text{A12})$$

The gap augmentation (A11) and total number of nans (A12) are correct only if the criteria defined in (A10) are satisfied. The moment that $\mathcal{S}_{\text{NaNs}}^{\text{ifo}} \subseteq \mathcal{S}_{\text{NaNs}}^{\mathbf{D}_{ij}(\text{ifo})}$ and $\mathcal{S}_{\text{NaNs}}^{\text{ifo}} \cap \mathcal{S}_{\text{NaNs}}^{\mathbf{D}_{ij}(\text{ifo})} \neq \emptyset$, the overall gap augmentation changes. In the next section, we discuss this in more detail.

Regime 2: Partial Merging

We are interested in finding the moment where the set defined in Eq. (A7) starts to merge with the original telemetry NaNs identified via Eq. (A1). This occurs when the last index of the set $\mathcal{S}_{\text{NaNs}}^{\text{ifo}}$ equals the first index of the set $\mathcal{S}_{\text{NaNs}}^{\mathbf{D}_{ij}(\text{ifo})}$, i.e.,

$$\mathcal{S}_{\text{NaNs}}^{\text{ifo}} \cap \mathcal{S}_{\text{NaNs}}^{\mathbf{D}_{ij}(\text{ifo})} = \{n_{\text{last}}^{\text{ifo}}\} = \{n_{\text{first}}^{\mathbf{D}_{ij}(\text{ifo})}\}. \quad (\text{A13})$$

The choice of Lagrange interpolant allowing this condition is

$$\mathcal{K} = 1 + 2N - 2N_{\text{NaNs}}^{\text{ifo}}. \quad (\text{A14})$$

Notice the transition point now depends on the choice of integer delay N and number of missing data samples in the telemetry variables $N_{\text{NaNs}}^{\text{ifo}}$. Similarly, another transition point is the first instance where the telemetry missing data set is entirely contained within the delayed missing data set. Mathematically,

$$\mathcal{S}_{\text{NaNs}}^{\text{ifo}} \cap \mathcal{S}_{\text{NaNs}}^{\mathbf{D}_{ij}(\text{ifo})} = \mathcal{S}_{\text{NaNs}}^{\text{ifo}}, \quad (\text{A15})$$

which will only happen when first telemetry NaN index of $\mathcal{S}_{\text{NaNs}}^{\text{ifo}}$ is equal to the first NaN index of $\mathcal{S}_{\text{NaNs}}^{\mathbf{D}_{ij}(\text{ifo})}$, i.e., when $\{n_{\text{first}}^{\mathbf{D}_{ij}(\text{ifo})}\} = \{n_{\text{first}}^{\text{ifo}}\}$, giving the solution

$$\mathcal{K} = 2N - 1. \quad (\text{A16})$$

The gap augmentation with respect to the telemetry variables is then given by the (inclusive) difference between the last element in $\mathcal{S}_{\text{NaNs}}^{\mathbf{D}_{ij}(\text{ifo})}$ and last element in $\mathcal{S}_{\text{NaNs}}^{\text{ifo}}$. A short calculation shows

$$N_{\text{NaNs}}^{\text{tel} \rightarrow \eta} = n_{\text{last}}^{\mathbf{D}_{ij}(\text{ifo})} - n_{\text{last}}^{\text{ifo}} + 1 = \frac{\mathcal{K} + 1}{2} + N. \quad (\text{A17})$$

which is valid under the condition that $\mathcal{K} \geq 1 + 2N - 2N_{\text{NaNs}}^{\text{ifo}}$ and $\mathcal{K} \leq 2N - 1$. The overall number of contiguous NaNs in Eq. (11) is given by

$$N_{\text{NaNs}}^{\eta} = \underbrace{N_{\text{NaNs}}^{\text{ifo}}}_{\text{Telemetry}} + \underbrace{\frac{\mathcal{K} + 1}{2} + N}_{\text{Gap Augmentation}}. \quad (\text{A18})$$

Our third and final regime for the gap augmentation occurs due to a complete merging between the two sets $\mathcal{S}_{\text{NaNs}}^{\mathbf{D}_{ij}(\text{ifo})}$ and $\mathcal{S}_{\text{NaNs}}^{\text{ifo}}$ and will be the focus of the next section.

Regime 3: Complete Merging

We now focus on the final (and simplest) case where the set of telemetry NaNs are contained within the set of delayed telemetry NaNs. In other words, when $\mathbf{D}_{ij} \text{ifo}_{ij}[n \in \mathcal{S}_{\text{NaNs}}^{\text{ifo}}] = \text{NaN}$ with $\mathcal{S}_{\text{NaNs}}^{\text{ifo}} \subseteq \mathcal{S}_{\text{NaNs}}^{\mathbf{D}_{ij}(\text{ifo})}$.

The statement that $\mathcal{S}_{\text{NaNs}}^{\text{ifo}} \subseteq \mathcal{S}_{\text{NaNs}}^{\mathbf{D}_{ij}(\text{ifo})}$ means that the gap segment in the telemetry variables is encapsulated within the missing data segment. From Eq. (A8), we understand that the *total number* of NaNs will be $\mathcal{K} + N_{\text{NaNs}}^{\text{tel}}$. This means that the overall gap augmentation for $\mathcal{K} > 2N - 1$ must be

$$N_{\text{NaNs}}^{\text{tel} \rightarrow \eta} = \mathcal{K}. \quad (\text{A19})$$

and overall number of missing data given by

$$N_{\text{NaNs}}^{\eta} = \underbrace{N_{\text{NaNs}}^{\text{ifo}}}_{\text{telemetry}} + \underbrace{\mathcal{K}}_{\text{Gap Augmentation}}. \quad (\text{A20})$$

Putting all of the pieces together with Gap Augmentations Eq. (A11), Eq. (A18), Eq. (A19) specific to the constraints Eq. (A14) and Eq. (A16) results in Eq. (21) defined in the main body of the text.

Appendix B: Gap Augmentation with Nested Delays

The gap augmentation formula leading to Eq. (21) holds when considering single delay $\mathbf{D}_{ij}$ on the ifo measurements. Deriving the gap augmentation where nested delays are utilized follows a near identical procedure discussed above. Consider, for example,

$$y[n] = x[n] + \mathbf{D}_{i_0 i_1 \dots i_{m-1}} x[n] \quad (\text{B1})$$

$$\approx x[n] + \sum_{k=1-p}^p x[n - N + k], \quad (\text{B2})$$

with integer part $N = \lfloor m \cdot \bar{L} / (c\Delta t) \rfloor$. Understand now that Eq. (B2) is in the same form as Eq. (A9) only with

a different integer part N given by the sum of delays of individual links at specific time-stamps. We remark here that this simple expression is only valid for equal

and constant arm-lengths. From near identical calculations that lead to Eq. (21), we can compute the gap augmentation induced in Eq. (B2).

$$N_{\text{NaNs}}^{x \rightarrow y} = f(\mathcal{K}, N_{\text{NaNs}}^x, N) = \begin{cases} \mathcal{K} + N_{\text{NaNs}}^x & \text{if } 1 < \mathcal{K} \leq 1 + 2N - 2N_{\text{NaNs}}^x \\ \frac{\mathcal{K}+1}{2} + N & 1 + 2N - 2N_{\text{NaNs}}^x \leq \mathcal{K} \leq 2N - 1 \\ \mathcal{K} & \mathcal{K} \geq 2N - 1. \end{cases} \quad (\text{B3})$$

For N_{NaNs}^x the number of contiguous nans in the sequence of x . Notice that the notation outlined in Eq. (B3) can be used to compute the gap augmentation in the $\eta_{ij}[n]$ variables, where $N_{\text{NaNs}}^{\eta \rightarrow \text{ifo}} = f(\mathcal{K}, N_{\text{NaNs}}^{\text{ifo}}, N)$.

The factorized expressions for the first and second generation TDI variables, derived from Eq. (23), Eq. (24) and Eq. (26), are written in an identical form to (B2). Equation (B3) will be used extensively in Section III C.

Appendix C: Multiple Gaps in Telemetry Variables

For multiple gaps in the telemetric data streams, we can construct the full set of G gaps.

$$\begin{aligned} \mathcal{S}_{\text{gaps}}^{\text{ifo}} &= \{(\mathcal{S}_{\text{NaNs}}^{\text{ifo}})^{(0)}, (\mathcal{S}_{\text{NaNs}}^{\text{ifo}})^{(1)}, \dots, (\mathcal{S}_{\text{NaNs}}^{\text{ifo}})^{(G-1)}\}, \\ &= \left\{ (\mathcal{S}_{\text{NaNs}}^{\text{ifo}})^{(i)} \right\}_{i=0}^{G-1} \end{aligned} \quad (\text{C1})$$

where each element of $(\mathcal{S}_{\text{NaNs}}^{\text{ifo}})^{(i)} \in \mathcal{S}_{\text{gaps}}^{\text{ifo}}$ is a gap consisting of a contiguous set of missing-data of the form identical to (A1)

$$(\mathcal{S}_{\text{NaNs}}^{\text{ifo}})^{(i)} = \{n_0^{(i)}, n_0^{(i)} + 1, \dots, n_0^{(i)} + (N_{\text{NaNs}}^{\text{ifo}})^{(i)} - 1\}. \quad (\text{C2})$$

Here $n_0^{(i)}$ stands for the first index of the i -th gap segment with $i \in \{0, \dots, G-1\}$ for G number of data gaps present. Each gap segment $(\mathcal{S}_{\text{NaNs}}^{\text{ifo}})^{(i)}$ contains at most $(N_{\text{NaNs}}^{\text{ifo}})^{(i)}$ and do not overlap, i.e.,

$$\bigcap_{i=0}^{G-1} (\mathcal{S}_{\text{NaNs}}^{\text{ifo}})^{(i)} = \emptyset. \quad (\text{C3})$$

We will now discuss how to compute the gap augmentation in the intermediary variables $\eta_{ij}[n]$ for the general case where we have arbitrary numbers of gapped segments ($G > 1$) each with potentially different missing data durations $(N_{\text{NaNs}}^{\text{ifo}})^{(i)} \neq (N_{\text{NaNs}}^{\text{ifo}})^{(j)}$.

For each gap segment representing missing data in the telemetry variables $(\mathcal{S}_{\text{NaNs}}^{\text{ifo}})^{(i)}$, we proceed in a similar fashion to Appendix A, Eq. (A6), by computing the first and last indices of the summation term in Eq. (A5)

where missing-data is identified for each gap segment $i \in \{0, \dots, G-1\}$

$$(n_{\text{first}}^{\mathbf{D}_{ij}(\text{ifo})})^{(i)} = n_0^{(i)} + N - p \quad (\text{C4})$$

$$(n_{\text{last}}^{\mathbf{D}_{ij}(\text{ifo})})^{(i)} = n_0^{(i)} + (N_{\text{NaNs}}^{(\text{ifo}, i)} - 1) + N - (1 - p). \quad (\text{C5})$$

With these indices we can then construct the masked data sets for $i \in \{0, \dots, G-1\}$

$$\mathcal{S}_{\text{gaps}}^{\mathbf{D}_{ij}(\text{ifo})} = \left\{ (\mathcal{S}_{\text{NaNs}}^{\mathbf{D}_{ij}(\text{ifo})})^{(i)} \right\}_{i=0}^{G-1} \quad (\text{C6})$$

with first (C4) and last (C5) indices of the missing data placement in the i^{th} missing data segment $(\mathcal{S}_{\text{NaNs}}^{\mathbf{D}_{ij}(\text{ifo})})^{(i)}$. To construct the overall gated segment in the intermediary variables $\eta_{ij}[n]$, corresponding to Eq. (A9), we merge the set $(\mathcal{S}_{\text{NaNs}}^{\mathbf{D}_{ij}(\text{ifo})})^{(i)}$ (in other words, take the union) with the original telemetry set $(\mathcal{S}_{\text{NaNs}}^{\text{ifo}})^{(i)}$ giving the missing data placements in gap i

$$(\mathcal{S}_{\text{NaNs}}^{\eta})^{(i)} = (\mathcal{S}_{\text{NaNs}}^{\mathbf{D}_{ij}(\text{ifo})})^{(i)} \cup (\mathcal{S}_{\text{NaNs}}^{\text{ifo}})^{(i)} \quad (\text{C7})$$

We now arrive at the complete data set representing the indices of NaNs present in the intermediary variables

$$\mathcal{S}_{\text{gaps}}^{\eta} = \{(\mathcal{S}_{\text{NaNs}}^{\eta})^{(0)}, (\mathcal{S}_{\text{NaNs}}^{\eta})^{(1)}, \dots, (\mathcal{S}_{\text{NaNs}}^{\eta})^{(G-1)}\}. \quad (\text{C8})$$

The overall gap augmentation is then calculated by computing the difference of missing data values between $\mathcal{S}_{\text{gaps}}^{\eta}$ and $\mathcal{S}_{\text{gaps}}^{\text{ifo}}$.

Repeated application of the procedure allows one to construct the gapped sets for the Michelson-like TDI variables, $\mathcal{S}_{\text{gaps}}^{X_1}$ and $\mathcal{S}_{\text{gaps}}^{X_2}$ for arbitrary numbers of gaps with varying durations. A very similar approach to Sec.(III C) is followed, where one computes the gapped data sets for the $a_{1,2}$, $r_{1,2}$ and $q_{1,2}$ variables defined by Equations (23), (24) and (26). The only considerable change is the overall integer delay that must be considered when moving from masked data sets $a_{1,2}$, $r_{1,2}$ and $q_{1,2}$. Due to the extra delays, we must modify the first and last NaN indices of the delayed data sets [(C4), (C5)] to have $N = \lfloor 2\bar{L}/(c\Delta t) \rfloor$ when augmenting from $a_{1,2} \rightarrow r_{1,2}$ and $N = \lfloor 4\bar{L}/(c\Delta t) \rfloor$ when augmenting from

$r_{1,2} \rightarrow q_{1,2}$. Following this procedure allows one to construct both first and second generation masked data sets for arbitrary families of gap types

$$\mathcal{S}_{\text{gaps}}^{X_1} = \{(\mathcal{S}_{\text{NaNs}}^{X_1})^{(0)}, (\mathcal{S}_{\text{NaNs}}^{X_1})^{(1)}, \dots, (\mathcal{S}_{\text{NaNs}}^{X_1})^{(G-1)}\}, \quad (\text{C9a})$$

$$\mathcal{S}_{\text{gaps}}^{X_2} = \{(\mathcal{S}_{\text{NaNs}}^{X_2})^{(0)}, (\mathcal{S}_{\text{NaNs}}^{X_2})^{(1)}, \dots, (\mathcal{S}_{\text{NaNs}}^{X_2})^{(G-1)}\}. \quad (\text{C9b})$$

where the overall gap augmentation and final total number of missing data elements can then be easily calculated.

The procedure above is a general method to compute the overall number of missing data through propagation of gated telemetry through TDI. We will use the formalism presented in this section when we analyse a full-scale four-year long LISA data set sampled at 4 Hz in Sec.(IV C). Such simulations are too memory intensive and expensive to be generated using LISA simulation software such as `lisainstrument` and `pyTDI`.

Our discussion in this appendix is general and is applicable to any family of gap types with any duration and rate. Our only requirement is that the missing data points are consistent across all IFO measurements.

Although the methods discussed here will always be cheaper than propagating corrupted telemetry data

through `pyTDI`, it still can be quite expensive to compute. Data sets that are four years in length sampled at 4 Hz are large to store and, although the operations discussed above are relatively simple, there is still a vast quantity of data that must be processed. A cheaper and still very accurate way to compute the gap augmentation is to simply compute the gap augmentation in the various TDI variables using the analytical formulas (35). Given the list of missing-data telemetry sets (C2), one can compute Eq. (35)

and sum the total amount of missing-data predicted from each segment.

This is a very cheap way to estimate the number of missing data points in the overall TDI expressions with the only caveat being a potential overestimation of the number of missing data values. Such scientific claims should then be understood as conservative. Making reference to Fig. 2, we see that for a fixed order of interpolate $\mathcal{K} = 45$, the maximum loss of data is approximately ~ 90 s. This means that if there are multiple (telemetry-based) gaps contained within a 90 second interval, then the cheap gap augmentation procedure outlined above will not account for the intersecting data gaps as a result of widened edges. Using this cheap method leads to a possibility that we will slightly overestimate the number of missing data elements, which is ultimately conservative at heart.

-
- [1] M. Colpi *et al.* (LISA), LISA Definition Study Report (2024), [arXiv:2402.07571 \[astro-ph.CO\]](#).
 - [2] P. Amaro-Seoane *et al.* (LISA), Laser Interferometer Space Antenna (2017), [arXiv:1702.00786 \[astro-ph.IM\]](#).
 - [3] M. Colpi *et al.*, (2019), [arXiv:1903.06867 \[astro-ph.GA\]](#).
 - [4] C. P. L. Berry, S. A. Hughes, C. F. Sopuerta, A. J. K. Chua, A. Heffernan, K. Holley-Bockelmann, D. P. Mihaylov, M. C. Miller, and A. Sesana, *The unique potential of extreme mass-ratio inspirals for gravitational-wave astronomy* (2019), [arXiv:1903.03686 \[astro-ph.HE\]](#).
 - [5] M. Tinto and S. V. Dhurandhar, *Living Reviews in Relativity* **17**, 6 (2014).
 - [6] J. Aasi and others (LIGO Scientific Collaboration), *Classical and Quantum Gravity* **32**, 074001 (2015).
 - [7] B. P. Abbott, others (LIGO Scientific, and V. Collaborations), *Phys. Rev. Lett.* **116**, 061102 (2016).
 - [8] M. Colpi, K. Danzmann, M. Hewitson, K. Holley-Bockelmann, P. Jetzer, G. Nelemans, A. Petiteau, D. Shoemaker, C. Sopuerta, R. Stebbins, N. Tanvir, H. Ward, W. J. Weber, I. Thorpe, A. Dauriskikh, A. Deep, I. F. Núñez, C. G. Marirrodiga, M. Gehler, J.-P. Halain, O. Jennrich, U. Lammers, J. Larrañaga, M. Lieser, N. Lützgendorf, W. Martens, L. Mondin, A. P. Niño, P. Amaro-Seoane, M. A. Sedda, P. Auclair, S. Babak, Q. Baghi, V. Baibhav, T. Baker, J.-B. Bayle, C. Berry, E. Berti, G. Boileau, M. Bonetti, R. Brito, R. Buscicchio, G. Calcagni, P. R. Capelo, C. Caprini, A. Caputo, E. Castelli, H.-Y. Chen, X. Chen, A. Chua, G. Davies, A. Derdzinski, V. F. Domcke, D. Doneva, I. Dvorkin, J. M. Ezquiaga, J. Gair, Z. Haiman, I. Harry, O. Hartwig, A. Hees, A. Heffernan, S. Husa, D. Izquierdo, N. Karnesis, A. Klein, V. Korol, N. Korsakova, T. Kupfer, D. Laghi, A. Lamberts, S. Larson, M. L. Jeune, M. Lewicki, T. Littenberg, E. Mudge, A. Mangiagli, S. Marsat, I. M. Vilchez, A. Maselli, J. Mathews, M. van de Meent, M. Muratore, G. Nardini, P. Pani, M. Peloso, M. Pieroni, A. Pound, H. Quelquejay-Leclerc, A. Ricciardone, E. M. Rossi, A. Sartirana, E. Savalle, L. Sberna, A. Sesana, D. Shoemaker, J. Slutsky, T. Sotiriou, L. Speri, M. Staab, D. Steer, N. Tamanini, G. Tasinato, J. Torrado, A. Torres-Orjuela, A. Toubiana, M. Vallisneri, A. Vecchio, M. Volonteri, K. Yagi, and L. Zwick, *Lisa definition study report* (2024), [arXiv:2402.07571 \[astro-ph.CO\]](#).
 - [9] H. Audley *et al.*, *arXiv e-prints* (2017), [arXiv:1702.00786 \[astro-ph.IM\]](#).
 - [10] O. Hartwig, J.-B. Bayle, A. Petiteau, *et al.*, *Classical and Quantum Gravity* **38**, 195021 (2021).
 - [11] M. Staab, O. Hartwig, and J.-B. Bayle, *arXiv e-prints* (2023), [arXiv:2312.11064 \[astro-ph.IM\]](#).
 - [12] J. N. Reinhardt, P. Euringer, O. Hartwig, K. Yamamoto, and K. Danzmann, *Physical Review D* **110**, 082005 (2024).

- [13] J.-B. Bayle, A. Petiteau, and H. Halloin, [arXiv e-prints \(2018\)](#), [arXiv:1811.01575 \[gr-qc\]](#).
- [14] M. Tinto, G. de Vine, J. C. Livas, K. McKenzie, and D. A. Shaddock, [Phys. Rev. D](#) **102**, 022003 (2020).
- [15] M.-S. Hartig, S. Schuster, and G. Wanner, [J. Opt.](#) **24**, 065601 (2022).
- [16] S. Paczkowski, R. Giusteri, M. Hewitson, N. Karnesis, E. D. Fitzsimons, G. Wanner, and G. Heinzel, [Phys. Rev. D](#) **106**, 042005 (2022).
- [17] M.-S. Hartig, S. Paczkowski, M. Hewitson, G. Heinzel, and G. Wanner, [Physical Review D](#) **111**, 10.1103/physrevd.111.043048 (2025).
- [18] J.-B. Bayle and O. Hartwig, [Phys. Rev. D](#) **107**, 083019 (2023).
- [19] European Space Agency (ESA), [Science performance requirements document for lisa](#) (2022), IISA Science Requirements Document (SciRD).
- [20] E. Castelli, Q. Baghi, J. G. Baker, J. Slutsky, J. Bobin, N. Karnesis, A. Petiteau, O. Sauter, P. Wass, and W. J. Weber, [Class. Quant. Grav.](#) **42**, 065018 (2025), [arXiv:2411.13402 \[gr-qc\]](#).
- [21] O. Burke, S. Marsat, J. R. Gair, and M. L. Katz, [Phys. Rev. D](#) **111**, 124053 (2025), [arXiv:2502.17426 \[gr-qc\]](#).
- [22] R. Mao, J. E. Lee, and M. C. Edwards, [Phys. Rev. D](#) **111**, 024067 (2025), [arXiv:2410.05571 \[gr-qc\]](#).
- [23] K. Dey, N. Karnesis, A. Toubiana, E. Barausse, N. Korsakova, Q. Baghi, and S. Basak, [Phys. Rev. D](#) **104**, 044035 (2021), [arXiv:2104.12646 \[gr-qc\]](#).
- [24] Q. Baghi, I. Thorpe, J. Slutsky, J. Baker, T. Dal Canton, N. Korsakova, and N. Karnesis, [Phys. Rev. D](#) **100**, 022003 (2019), [arXiv:1907.04747 \[gr-qc\]](#).
- [25] J. Carre and E. K. Porter, [arXiv:1010.1641 \[gr-qc\]](#) (2010).
- [26] L. Wang, H.-Y. Chen, X. Lyu, E.-K. Li, and Y.-M. Hu, [Eur. Phys. J. C](#) **85**, 120 (2025), [arXiv:2405.14274 \[gr-qc\]](#).
- [27] T. B. Littenberg and N. J. Cornish, [Phys. Rev. D](#) **107**, 063004 (2023), [arXiv:2301.03673 \[gr-qc\]](#).
- [28] M. L. Katz, N. Karnesis, N. Korsakova, J. R. Gair, and N. Stergioulas, [Phys. Rev. D](#) **111**, 024060 (2025), [arXiv:2405.04690 \[gr-qc\]](#).
- [29] S. Deng, S. Babak, M. Le Jeune, S. Marsat, É. Plagnol, and A. Sartirana, [Physical Review D](#) **111**, 103014 (2025).
- [30] S. H. Strub, L. Ferraioli, C. Schmeltzbach, S. C. Stähler, and D. Giardini, [Phys. Rev. D](#) **110**, 024005 (2024).
- [31] P. Whittle, [Journal of the Royal Statistical Society: Series B \(Statistical Methodology\)](#) **19**, 38 (1957).
- [32] C. Talbot, E. Thrane, S. Biscoveanu, and R. Smith, [Phys. Rev. Res.](#) **3**, 043049 (2021), [arXiv:2106.13785 \[astro-ph.IM\]](#).
- [33] O. Edy, A. Lundgren, and L. K. Nuttall, [Phys. Rev. D](#) **103**, 124061 (2021), [arXiv:2101.07743 \[astro-ph.IM\]](#).
- [34] J.-B. Bayle, O. Hartwig, and M. Staab, [Lisa instrument](#), Zenodo (2023).
- [35] M. Staab, J.-B. Bayle, and O. Hartwig, [Pytdi \(1.2.1\)](#) (2022).
- [36] M.-S. Hartig, S. Schuster, and G. Wanner, [Journal of Optics](#) **24**, 065601 (2022).
- [37] M. Chwalla, K. Danzmann, M. D. Álvarez, J. E. Delgado, G. Fernández Barranco, E. Fitzsimons, O. Gerberding, G. Heinzel, C. Killow, M. Lieser, M. Perreut-Lloyd, D. Robertson, J. Rohr, S. Schuster, T. Schwarze, M. Tröbs, G. Wanner, and H. Ward, [Phys. Rev. Appl.](#) **14**, 014030 (2020).
- [38] O. Hartwig and J.-B. Bayle, [Phys. Rev. D](#) **103**, 123027 (2021).
- [39] D. Q. Nam, Y. Lemiere, A. Petiteau, J.-B. Bayle, O. Hartwig, J. Martino, and M. Staab, [Tdi noises transfer functions for lisa](#) (2022), [arXiv:2211.02539 \[gr-qc\]](#).
- [40] M. Muratore, [Time delay interferometry for LISA science and instrument characterization](#), [Ph.D. thesis](#), Trento U. (2021).
- [41] M. Tinto, F. B. Estabrook, and J. W. Armstrong, [Phys. Rev. D](#) **65**, 082003 (2002).
- [42] O. Hartwig, J.-B. Bayle, M. Staab, A. Hees, M. Lilley, and P. Wolf, [Phys. Rev. D](#) **105**, 122008 (2022).
- [43] M. Muratore, D. Vetrugno, and S. Vitale, [Class. Quant. Grav.](#) **37**, 185019 (2020), [arXiv:2001.11221 \[astro-ph.IM\]](#).
- [44] M. Vallisneri, [Phys. Rev. D](#) **72**, 042003 (2005), [Erratum: [Phys.Rev.D](#) 76, 109903 (2007)], [arXiv:gr-qc/0504145](#).
- [45] M. Le Jeune and S. Babak, [10.5281/zenodo.7132178](#) (2022).
- [46] M. Le Jeune and S. Babak, [10.5281/zenodo.7132178](#) (2022).
- [47] M. Staab, M. Lilley, J.-B. Bayle, and O. Hartwig, [Phys. Rev. D](#) **109**, 043040 (2024).
- [48] M. Staab, J.-B. Bayle, O. Hartwig, A. Hees, M. Lilley, G. Woan, and P. Wolf, [Class. Quant. Grav.](#) **42**, 115002 (2025), [arXiv:2412.14884 \[astro-ph.IM\]](#).
- [49] O. Hartwig, [Instrumental modelling and noise reduction algorithms for the Laser Interferometer Space Antenna](#), [Ph.D. thesis](#), Leibniz U., Hannover (2021).
- [50] O. Hartwig, J.-B. Bayle, M. Staab, A. Hees, M. Lilley, and P. Wolf, [Phys. Rev. D](#) **105**, 122008 (2022), [arXiv:2202.01124 \[gr-qc\]](#).
- [51] W. Martens and E. Joffre, [The Journal of the Astronomical Sciences](#) **68**, 402–443 (2021).
- [52] M. Staab, J.-B. Bayle, O. Hartwig, A. Hees, M. Lilley, G. Woan, and P. Wolf, [Classical and Quantum Gravity](#) **42**, 115002 (2025).
- [53] J.-B. Bayle, M. Lilley, A. Petiteau, and H. Halloin, [Phys. Rev. D](#) **99**, 084023 (2019).
- [54] N. J. Cornish, [Physical Review D](#) **102**, 124038 (2020), [arXiv:2009.00043 \[gr-qc\]](#).
- [55] N. Pearson and N. J. Cornish, [Handling data gaps for the next generation of gravitational-wave observatories](#) (2025), [arXiv:2509.05479 \[gr-qc\]](#).
- [56] M. Muratore, D. Vetrugno, S. Vitale, and O. Hartwig, [Phys. Rev. D](#) **105**, 023009 (2022).
- [57] O. Hartwig, J.-B. Bayle, M. Staab, A. Hees, M. Lilley, and P. Wolf, [Physical Review D](#) **105**, 10.1103/physrevd.105.122008 (2022).
- [58] N. Houba, J.-B. Bayle, and M. Vallisneri, [Robust bayesian inference with gapped lisa data using all-in-one tdi- \$\infty\$](#) (2024), [arXiv:2412.20793 \[astro-ph.IM\]](#).
- [59] J.-B. Bayle, M. Staab, S. P. Francis, E. R. Rees, R. Spero, and G. Heinzel, [Class. Quant. Grav.](#) **42**, 155009 (2025), [arXiv:2507.19267 \[gr-qc\]](#).
- [60] J. N. Reinhardt, M. Staab, K. Yamamoto, J.-B. Bayle,

- A. Hees, O. Hartwig, K. Wiesner, S. Shah, and G. Heinzl, [Phys. Rev. D](#) **109**, 022004 (2024).
- [61] J.-B. Bayle, O. Burke, E. Castelli, and N. Korsakova, [Lisa glitch](#) (2022).