

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, United Kingdom*

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



(Received 4 November 2025; revised 21 December 2025; accepted 13 January 2026; published 12 March 2026)

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 it 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 approximately 84%, we find that frequent, short-duration gaps (e.g., a total of 1000 per year, each of which has a duration of approximately 100 s) could result in an additional loss in the TDI variables of about one day per year, corresponding to a reduction of approximately 0.3% in duty cycle. This amounts to a loss of approximately one day of LISA data suitable for the global fit per year.

DOI: [10.1103/xpvf-syrw](https://doi.org/10.1103/xpvf-syrw)

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 at frequencies between 0.1 mHz and 1 Hz [1]. Unlike ground-based detectors, LISA is expected to be signal dominated, observing a multitude of GWs that overlap in both time and frequency [2,3]. Current GW-inference techniques for LISA are not designed to operate directly on the raw phasemeter data; instead, they take preprocessed data from an initial noise-reduction pipeline that transforms the raw phasemeter measurements into calibrated data streams suitable for analysis [4]. 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 8 to 9 orders of magnitude in amplitude. In contrast, ground-based GW detectors such as LIGO, VIRGO, and KAGRA

are able to suppress laser noise almost perfectly by constructing near-equal-arm interferometers [5,6]. 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° to the ecliptic [1]. While the nominal arm lengths are approximately 2.5×10^6 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 [7–11].

Each LISA spacecraft (S/C) hosts two moving optical subassemblies (MOSAs), and each of these carries an optical bench (OB) equipped with three distinct interferometers (IFOs): (i) an intersatellite IFO (sci), (ii) a test-mass IFO (tmi), and (iii) a reference IFO (ref) [1]. The test-mass IFO measures subnanometer fluctuations in the displacement between the OB and its associated free-falling test mass, which provides a local inertial reference. The intersatellite IFO interferes the incoming light from a remote spacecraft with locally generated laser light, measuring the relative optical path length between the distant OB and the local OB. Finally, the reference interferometer combines

*Contact author: ollie.burke@glasgow.ac.uk

†Contact author: martina.muratore@aei.mpg.de

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

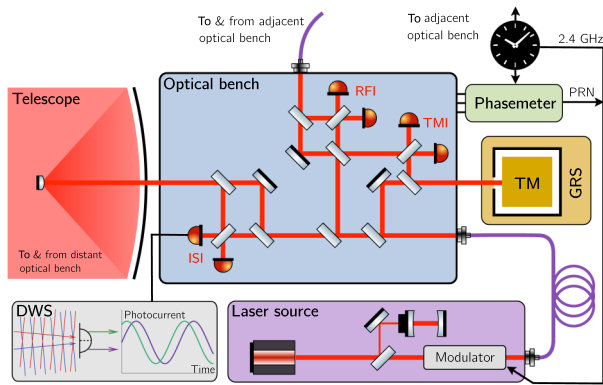


FIG. 1. LISA’s measurement system, adapted from [16] with support from O. Hartwig. This schematic shows how each spacecraft uses three main interferometers (IFOs) to detect GWs. The interspacecraft IFO (ISI) captures the GW signal, buried within much louder noise, primarily from laser frequency fluctuations. The other two IFOs, the reference and test-mass IFOs (RFI and TMI, respectively), help remove noise due to spacecraft motion and reduce the number of lasers from six to three. Additionally, clock sidebands and pseudorandom noise codes correct for timing and laser noise, while the differential wavefront sensing measures beam angles to minimize the effects of angular jitter.

the laser from the adjacent OB with the local laser. Its main purpose is to simplify the LISA measurement chain by comparing the six lasers via optical fibers, effectively reducing them to three [12]. 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 interspacecraft 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 these are used to minimize the effects of angular jitter [13–15]. All these measurements are sampled at 4 Hz before transmission to Earth, and they are summarized in Fig. 1.

Given the complexity of the LISA instrument and operations [17], data artifacts are to be expected [18,19]. Here, we will focus on the impact of missing data (that is, data gaps) potentially due to corrupted data that is removed having been rendered unsuitable for further analysis. The gap patterns 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 that several such events may occur annually, each lasting anywhere from 1 to 3 days.

Routine maintenance, as well as periods of partial or complete data loss, is 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 authors’ best knowledge, all existing studies on the impact of data gaps [18–24] have focused on signal recovery from postprocessed 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 that are 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-characterized datasets. The global fit for LISA aims to jointly recover both gravitational-wave sources and the underlying noise properties [25–28]. Accurate noise modeling is crucial since likelihood functions such as the Whittle likelihood [29] 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 [30,31] and noncirculant (Toeplitz-like) noise, making the Whittle likelihood only an approximation. Mischaracterizing the noise can bias parameter estimation [19], 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.

The remainder of this paper is organized as follows. In Sec. II, we give a brief discussion of TDI in Sec. II A and the Lagrange-interpolation methods used to compute fractional delays in Sec. II B. Section III contains our main

analytical results. Specifically, in Sec. III A, we discuss our analytical treatment of missing data and gaps; 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 are presented 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 Secs. 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 postprocessing on missing data. Finally, throughout our work, we will follow the conventions of PYTDI and LISAINSTRUMENT [32,33].

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 the overall scope of this work. On this subject, we will not describe all instrumental noise sources that are both well-characterized and measurable. This includes, for example, the optical tilt-to-length coupling induced by spacecraft motion and clock noise (see Refs. [34–36]). Residual noise sources refers to additional sources of noise that cannot be mitigated by the initial noise-reduction pipeline. We can group these residual noise sources into two main types: a test-mass (TM) acceleration noise and a general optical metrology system (OMS) noise contribution associated with each individual interspacecraft link. We refer the reader to Ref. [37] for extensions involving multiple OMS noise components).

For our analysis, we will consider only three IFO measurements [12] (see Fig. 1 for reference) for each optical bench:

- (a) sci_{ij} , the science IFO measuring the optical path between the laser hosted in the optical bench of the distant S/C j and the local laser i ;
- (b) ref_{ij} , the reference IFO interfering the local laser i with the laser in the adjacent optical bench j ;
- (c) tmi_{ij} , the test-mass IFO interfering the laser in the adjacent optical bench j with the local laser i once it has bounced off the test mass.

Note that here, the indices ij refer to signals emitted by S/C j and received by S/C i . In this work, we formulate the TDI observables in terms of total-frequency rather

than phase measurements [38]. In this representation, time delays experience Doppler effects and are therefore implemented via Doppler-delay operators that account for the light-travel time between spacecraft [39].

A. Construction of time-delay-interferometry variables

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

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

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

For a generic time series $x(t)$, the Doppler delay is applied as [40,41]

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

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 S/C j and S/C i with arm length L_{ij} , in which $d_{ij}(t) = L_{ij}(t)/c$.

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

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

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

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 nested delays, we define chained delay operators [40]:

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

which apply delays recursively. For instance,

$$\begin{aligned} \dot{\mathbf{D}}_{ijk}x(t) &= \dot{\mathbf{D}}_{ij}\dot{\mathbf{D}}_{jk}x(t) = [1 - \dot{d}_{ij}(t)][1 - \dot{d}_{jk}(t - d_{ij}(t))]x \\ &\times (t - d_{ij}(t) - d_{jk}(t - d_{ij}(t))), \end{aligned}$$

demonstrating the noncommutative 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.

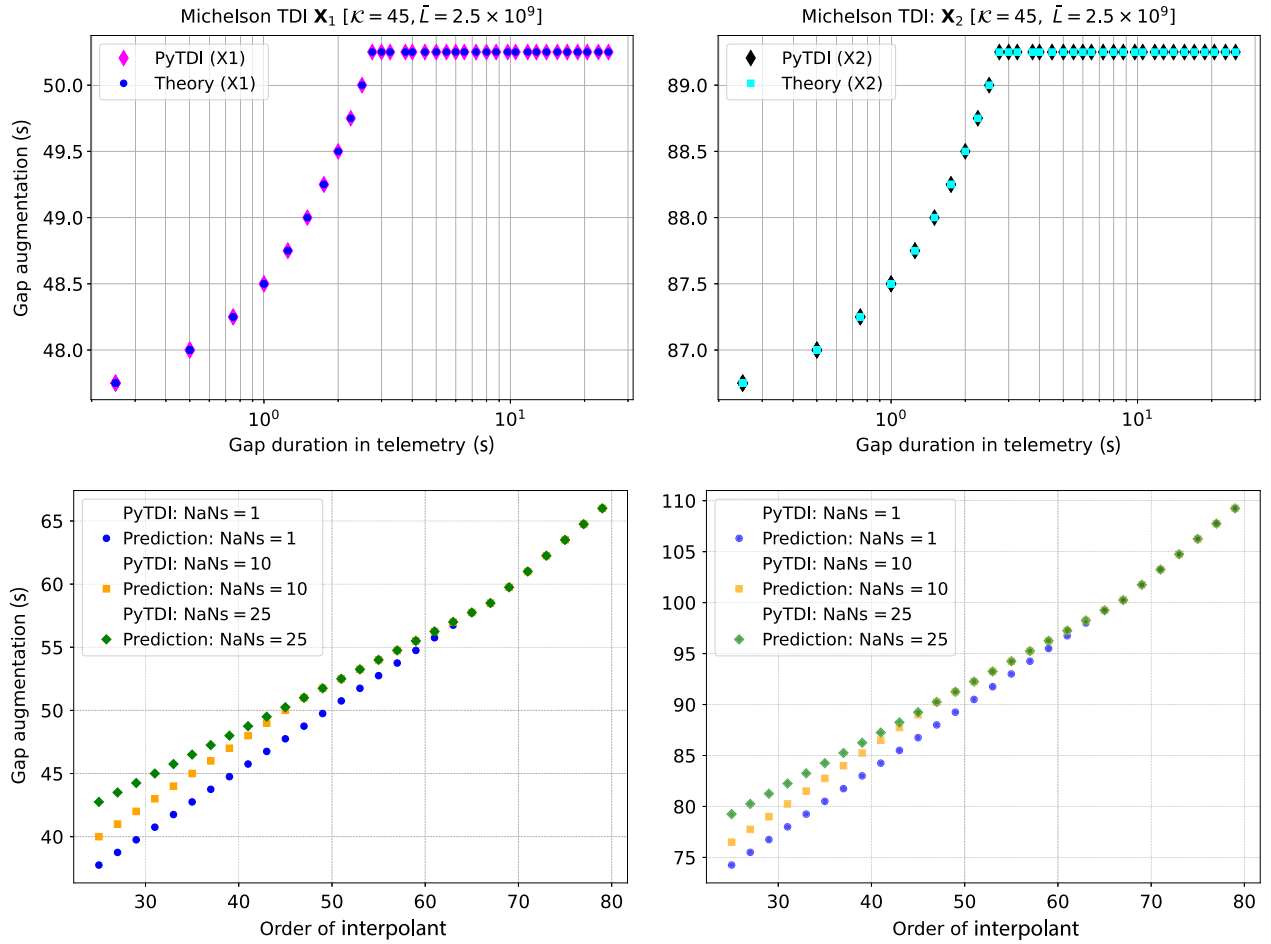


FIG. 2. Top-left and top-right plots: for first- and second-generation TDI variables [Eq. (25)], we compute the gap augmentation with $\mathcal{K} = 45$ for all Lagrange interpolants as a function of the gap duration. The purple diamonds and blue dots in the top-left plot compare the numerical and theoretical gap augmentation in the first-generation TDI variables using LISAINSTRUMENT and PYTDL and Eq. (35a). The second plot is the same but for second-generation TDI variables using PYTDL, and Eq. (35b), respectively. Bottom-left and bottom-right plots: here, we compare the gap augmentation using LISAINSTRUMENT and PYTDL against our analytical result of Eq. (35a) (left plot) and of Eq. (35b) (right plot) as a function of the order of the 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 the interpolant [see Eq. (21) for similar behavior 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 $N_{\text{NaNs}}^{\text{ifo}} = \{1, 10, 25\}$. A JUPYTER notebook with code to reproduce these plots is available from https://github.com/OllieBurke/gaps_tdi.

First- and second-generation TDI observables, which incorporate increasingly longer sequences of delays to accommodate time-dependent arm lengths, are given in factorized form [33] as

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

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

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 [42,43]; here, we will only discuss the Michelson TDI because it is the current baseline used for LISA data analysis [44].

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 $t_n = t_0 + n\Delta t$,

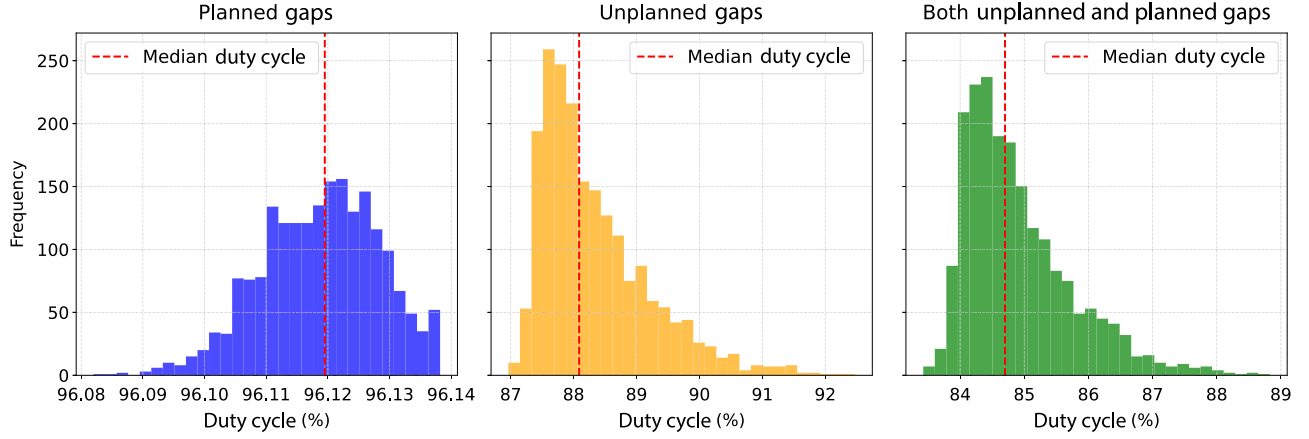


FIG. 3. Distributions of telemetry duty cycles under different gap strategies. The left-hand, center, and right-hand plots show the results for planned, unplanned, and both planned and unplanned gaps, respectively, using the rates in Table I.

in which t_0 is the initial starting time and $\Delta t = t_{n+1} - t_n$ is 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 $\tilde{\mathbf{D}}$ since the term $\tilde{d}_{ij}$ 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* [45,46]. Lagrange interpolation has a maximally flat transfer function at $f = 0$, which implies that this interpolation method is very accurate for low frequencies, where LISA will operate. A more efficient—and numerically more stable—interpolant was recently developed in Ref. [47], which avoids: (1) pathological behavior at higher frequencies; (2) an elevated number of terms (coefficients) in the interpolation scheme required to accurately compute the delay; and finally (3) prominent time-domain glitches and associated spectral leakage that occur when time-varying delays cross integer multiples of the sampling period. The interpolation method found in Ref. [47] could easily be incorporated into this work, but this would not change the results significantly.

Now, suppose the delay $d_{ij}(t) \in \mathbb{R}^+$. Then, following Ref. [48], we can always make the split

$$d_{ij}(t) = d_{ij}^{\text{int}}(t) + d_{ij}^{\text{frac}}(t), \quad (7)$$

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

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

$$N_{ij}^{\text{frac}}[n] = d_{ij}[n]/\Delta t - N_{ij}^{\text{int}}[n]. \quad (9)$$

Notice that the integer part $N_{ij}^{\text{int}}[n]$ could be (1) time varying due to the orbital dynamics and (2) unequal but constant between individual links (i, j) , in case of constant but not equal arm lengths. Figure 3.5 of Ref. [49] depicts a different scenario. If we are to assume the individual links to be equal and constant, then $N_{ij}^{\text{int}}[n] = N$, i.e., a constant integer delay that is common across all arm links. Similarly, for a static and equal-arm-length LISA configuration, the fractional quantity $N_{ij}^{\text{frac}}[n] = N^{\text{frac}} \in [0, 1)$ will also be constant across all single links.

In terms of continuous IFO measurements, we have

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

which, in discrete notation, is given by

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

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

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

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 Ref. [50], for simplicity, let $A_1, A_2, \dots, A_n$ represent potential S/C index pairings. Note that

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

with total time delay

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

which 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 nor constant, the delays are smooth functions with very-low-frequency content. For this reason, the recursive computation defined via Eq. (15) can be performed with very low interpolation orders that still result in highly accurate interpolated measurements [51].

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 $\bar{L} = L_{ij} = 2.5 \times 10^9$ 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 conclusions, 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 this 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

$$\begin{aligned} N &= \lfloor m \times \bar{d} \rfloor \quad (\text{integer part}), \\ \epsilon &= m \times \bar{d} - N \quad (\text{fractional part}), \end{aligned}$$

where m is the number of delay operators and $\bar{d}$ is the average light-travel time ($\bar{d} = 8.34$ s [52]) defined between any two links.

III. EFFECT 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 [19, 24]. This amounts to a multiplication of the original time series by a gating function that is 0 for $t \in T_{\text{gap}}$ and 1 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. [19]). After TDI has been performed, the *global-fit* analyses should, in theory, be invariant to whether you apply windows or masks to the 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 a failure to cancel laser noise at the boundaries of the gaps. Imputing the scientific telemetry data with zero values is particularly dangerous since this would lead to spurious noise artifacts at the gap boundaries in the final TDI streams. These pathological artifacts could then result in biased parameter-inference schemes conducted throughout global-fit-based pipelines. A much cleaner and more elegant method is to assume complete ignorance of missing data via imputation with unrepresentable and undefined floating points: NaNs [53].

For the discussion that follows, we will define

(a) Valid data: the contiguous set of defined numerical data that are not corrupted by NaNs.

(b) Missing data: specific elements of data that are undefined and cannot be represented by floating-point numbers. We will treat these numerically with NaNs.

(c) 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).

(d) Gap augmentation: the additional loss of scientific data (gap widening) after applying series of delays, e.g., computing intermediary variables or TDI Michelson-like variables.

To simplify the discussion, we will assume that each of the IFOs has 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 $N_{\text{NaNs}}^{\text{ifo}}$. Now, defining the set $\mathcal{S}_{\text{NaNs}}^{\text{ifo}}$ as the set [54] of indices $n \in \mathcal{S}_{\text{NaNs}}^{\text{ifo}}$ such that the IFO time series has contiguous missing data, we have

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

such that

$$\begin{aligned} \text{ifo}_{ij}[n \in \mathcal{S}_{\text{NaNs}}^{\text{ifo}}] &= \text{sci}_{ij}[n \in \mathcal{S}_{\text{NaNs}}^{\text{ifo}}] \\ &= \text{tmi}_{ij}[n \in \mathcal{S}_{\text{NaNs}}^{\text{ifo}}] \\ &= \text{ref}_{ij}[n \in \mathcal{S}_{\text{NaNs}}^{\text{ifo}}] \\ &= \text{NaNs}. \end{aligned} \quad (18)$$

The string of equalities in Eq. (18) should be interpreted as a loss of communication between the instrument and ground segment; however, this is only one of many possible occurrences of data loss. For instance, we may find that the IFOs on board S/C 1 exhibit significant glitchy behavior that results in total data loss, whereas the IFOs on S/C 2 and S/C 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 IFO 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 values exists in the interferometric measurements.

Recall from Eq. (18) that each of the IFO measurements exhibits 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), for static instrument configurations, we must have $\eta_{ij}[n \in \mathcal{S}_{\text{NaNs}}^{\eta}] = \text{NaNs}$ 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 rewrite Eqs. (1) and (3) to focus on the individual components hampered by missing data. The goal

here is to understand the gap augmentation of the following equation:

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

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 values (NaNs) in our calculations. This will simplify the expressions that feed into the TDI variables. In the discrete domain, we need to compute

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

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

Following Eqs. (17) and (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 $N = \lceil \bar{L}/(c\Delta t) \rceil$. As derived in Appendix A, the overall gap augmentation in the intermediary variables [Eq. (20)] is given by the piecewise function

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

for $\mathcal{K}$ of 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

$$N_{\text{NaNs}}^{\eta} = N_{\text{NaNs}}^{\text{ifo}} + N_{\text{NaNs}}^{\text{ifo} \rightarrow \eta}. \quad (22)$$

We remark that, in general, if one were to use the more numerically stable (and efficient) interpolant found in Ref. [47], the result given in Eq. (21) will remain unchanged. The interpolation scheme in Ref. [47] requires far fewer coefficients (22) than the Lagrange interpolants used throughout this work (with default Lagrange interpolation order set to 45). This implies that there would be less gap augmentation since we use fewer 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% peak to peak over the course of the mission, corresponding to deviations of about approximately $\pm 1.5\%$ around the mean arm length [52]. 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 the 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 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 interpolant of order $\mathcal{K}$ and an original gap length $N_{\text{NaNs}}^{\text{ifo}}$ that satisfy the intermediary regime set in Eq. (21), the regime in which the integer time shift determines the size of the gap widening.

To conclude, we strongly believe that our results (and conclusions) in Sec. 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 that we have a single set of contiguous missing data in the telemetry variables, resulting in a single gap. We remind the reader that since the telemetry variables all have the same missing data at the same time stamps, we understand that each of the six 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 generations of TDI variables $\mathbf{X}_1$ and $\mathbf{X}_2$, defined in Eqs. (5) and (6), is given in Ref. [55]. First, we can define the quantities

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

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

which can be understood as a single photon-path round trip. Then, defining a 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

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

Note that to build $\mathbf{X}_1$, we require only four interpolation schemes rather than the six required for the nonfactorized first-generation TDI expressions [38,45].

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 the 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 for 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 Eqs. (23), (24), and (26) is of a similar form to Eq. (28) provided that each has 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

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

for n delay operators for each of Eqs. (23), (24), and (26), and N_{NaNs}^x , the number of missing data values 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 Eqs. (23a) and (23b). Since each IFO has missing data at precisely the same time stamps, this implies that the gap augmentation in a_1 should be identical to that in 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 the overall numbers of missing data values in a_1 and a_2 given by

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

in which N_{NaNs}^η is 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 the total number of NaNs present in the r_1 and r_2 variables being

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

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 (33) can be used to find out the total number of missing data values in both first- and second-generation TDI variables. Combining each of these equations leads to the main result of our paper:

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

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

Equations (35a) and (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 Eq. (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 dataset with a realistic missing-data and gap scenario in Sec. IV C. We make use of our analytical results because with the current version of LISAINSTRUMENT, 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 setup

In this section, we will make use of LISAINSTRUMENT [32] to generate a small-scale LISA-based simulation for a static constellation with equal arm lengths set to $\bar{L} = 2.5 \times 10^9$ m. We will consider a sampling rate $f_s = 4$ Hz and disable all noise sources except for laser, TM, and OMS noise. The overall simulation will be around 8.3 h 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 LISAGLITCH [56] to generate 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 (sci_{ij} , tmi_{ij} , and ref_{ij}) by the masking function w . We do not apply a mask to the IFO sidebands and MPRs since we do not account for the effects of clock noise or evolving LISA arm lengths.

These masked telemetry data are fed into PYTDI [33] to compute the intermediary variables and resultant first- and second-generation TDI $\mathbf{X}$ variables. For the first analysis, where we compute the gap augmentation as a function of gap duration, we fix the order of the interpolants for both the intermediary and Michelson variables to $\mathcal{K} = 45$. An order $\mathcal{K} = 45$ is standard within current noise-mitigation pipelines since this choice balances computational efficiency and interpolation noise. See Ref. [55] for more information. In the second analysis, we will investigate the gap augmentation for three gap-duration scenarios as a function of the interpolant order $\mathcal{K} \in [25, 80]$. For all investigations, 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 $\mathcal{A}$, the order of the Lagrange interpolant used for the shift [57]. 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 results are given in Fig. 2. In all four panels of this figure, we show that our analytical result of Eq. (35) is consistent with the numerical gap augmentation obtained 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. In the lower panels, we also analyze three scenarios, with 1, 10, and 25 NaNs, and we study the gap augmentation with respect to the order of interpolation.

In the upper 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 is equal to

- (a) 50.25 s for first-generation variables,
- (b) 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. Intuitively, this can be understood from (1) the overall number of delays required to compute the Michelson-like variables and (2) the gap-augmentation regime one enters that is subject to the gap duration, integer time shift, and order of Lagrange interpolant. For gap durations of approximately 2.5 s, we find that the third gap regime in Eq. (21) is the primary consideration, where the gap augmentation is driven only by the order of the interpolant. Counting the number of delay operations between Eqs. (6) and (23), we find a total of eight. Each delay incorporates a gap augmentation equivalent to $\mathcal{K}$. Hence, the overall gap augmentation is approximately $8 \times \mathcal{K} \times \Delta t = 8 \times 45 \times 0.25 = 90$ s. Notice further that the third gap augmentation is independent of the gap duration. Small deviations in this number come from the earlier stages of the noise-mitigation pipelines (computing η , a , etc.) and from the potential for new gaps resulting from the interpolant merging (second regime).

Frequent and short gaps may be the most concerning. A loss of a single packet of data (approximately 0.25 s) can lead to an additional loss of approximately 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 around a few hours, while PAAM-related gaps are typically around 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 approximately 90 s, this is negligible compared with scheduled long gaps (e.g., estimation of tilt-to-length coupling coefficients), which can last up to approximately 40 h. Although any additional loss of data is undesirable, its relative impact in this case will be minimal. In summary, when comparing losses from TDI operations with 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 year⁻¹, approximately 100 s each), making the TDI-induced loss comparable to the intrinsic loss. Conversely,

for long gaps, which are rare (6 year⁻¹), the additional TDI loss is negligible compared with 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 at most by approximately 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 the overall delay expressed by

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

For multiple delays that determine the prefactor (n) in Eq. (36), the choice of $\bar{L} \in [\bar{L}_{\min}, \bar{L}_{\max}]$ will govern smaller or larger values of the integer time shift, where $\bar{L}_{\min}$ and $\bar{L}_{\max}$ can be estimated from the precomputed orbits given in Ref. [52]. The value of this integer time shift determines the regime of gap augmentation located in Eq. (B3). Note that the only regime in which 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 in our case, 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 or decrease (by only a few points) only if our order of interpolant, and the 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 , and 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 $\bar{L} \in [L_{\min}, L_{\max}]$, and our observations were as expected. For $\mathbf{X}_1$ and $\mathbf{X}_2$, we found that we underestimated the overall gap augmentation by around 1 and 3 points, respectively, if $\bar{L} = \bar{L}_{\max}$ or $\bar{L} = \bar{L}_{\min}$ for a short simulation of approximately 8.3 h. 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 provides a solid baseline for drawing scientific conclusions. 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 approximately 82% of usable data, the so-called *duty cycle* of the mission [1]. 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 data.

Table I, inspired by Ref. [18] 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. To the authors' knowledge, this table represents our best understanding of potential gap scenarios LISA may face. 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 Refs. [18,19,21].

We sample the planned data gaps at regular intervals following the rate (year^{-1}) given in Table I, with an approximately 10% random jitter applied to the gap duration (start time) of each subtype of gap. For the unplanned data gaps, the gap durations (start times) are sampled from an exponential distribution derived according to the rate (year^{-1}) given in Table I. The result is a binary mask with the integer value 1 representing valid data and NaNs representing missing data. The simulation techniques for the (un)planned binary masks have been drawn from Ref. [18].

In what follows, 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 planned gaps, one for unplanned gaps, and one with both planned and unplanned gaps—with rate parameters and durations given by the second and third columns of Table I.

Based on approximately 2000 seeded realizations of the four-year masking function, sampled at $\Delta t = 10$ s with the parameters specified in Table I, Table I summarizes the median number of masked segments, total hours lost, and duty cycles. The distributions of the telemetry duty cycles are given in Fig. 3. These results illustrate the following:

- (a) Unplanned data gaps account for the majority data loss for the instrument, accounting for a median of approximately 11.855%, computed considering an 88.145% duty cycle.
- (b) Planned data gaps only account for a median of approximately 3.888%, computed considering a 96.112% duty cycle.
- (c) Therefore, the overall (median) duty cycle is $100\% - (11.855 + 3.888)\% = 84.257\%$, which is greater than the quoted 82% nominal duty cycle [1,18].

This latter point means that we retain approximately 2.257% (35.61 days) of scientific data compared with the previously estimated (nominal) 82% duty cycle [1,18]. The nominal 82% duty cycle only accounts for the individual gaps affecting the LISA data rather than their combination [58], which is outlined in Table I. This would result in an increase of approximately 35.61 days to garner more science from GW observations. We remark here that these conclusions are based on a specific gap scenario, as outlined in Table I, with approximately 82% duty cycle computed via *nonoverlapping* gap segments.

TABLE I. Details of potential situations in which the spacecraft is affected by planned and unplanned data gaps. The first column represents the class of gaps, be they planned, unplanned, or a mixture of both. The second column represents the subtype of gap, with the third and fourth columns giving expected rates per year and durations (in hours). The fifth, sixth, and seventh columns respectively represent the median number of gap segments, total hours of missing data, and overall duty cycle for approximately 2000 seeded masking simulations sampled at $\Delta t = 10$ s with an 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.

Gap type	Subtype	Rate (year^{-1})	Duration (h)	(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 II. Summary of the effects of simulated gaps in telemetry and TDI variables. The first column represents the gap type, with gap parameters defined in Table 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 assuming a mission duration of four years.

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

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 Table I, at a sample rate of 4 Hz. To compute the overall number of missing data values in the telemetry variables, for arbitrary numbers of gaps and arbitrary durations, 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 presented in Table II.

1. Unplanned gaps

Here, 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 that of the planned gaps. The results are summarized in Table II. Here, 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 four-year observation period.

For unplanned data gaps, we lose out on approximately 3 h of data after applying TDI (see Table II), a small fraction when compared with the overall loss of data due to the unplanned data gaps. Since the order of the 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 in duration compared with the planned data gaps, the gap widening is constant and small, resulting in an augmentation (after TDI has been applied) of approximately 90 s per gap.

Given the family of unplanned gaps tabulated in Table I, when not accounting for the merging of the gaps for the unplanned dataset, we would expect approximately $4 \times (30 + 4 + 3) \times 89.25 \text{ s} \approx 0.15$ days of lost data, where we have considered the total rates, the gap augmentation of 89.25 s for TDI X_2 , and four years of observation. This

is a reasonable (but overestimated) approximation when compared with Table II, which 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 for the unplanned gaps, approximately 4.4852 days, which is approximately 35 times more data lost than for the unplanned data gaps. This is primarily due to the high-rate and short-duration data gaps with rate 1095 year^{-1} . Indeed, from a back-of-the-envelope calculation, we see that there should be $4 \times (6 + 26 + 1095) \times 89.25 \text{ s} \approx 4.66$ days of lost data. We see from Table II that our estimate is reasonable, with the true result for the total lost LISA data for the planned scheme being approximately 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 approximately 4 days 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 run through TDI-based noise-mitigation pipelines. Our results here support reports in the literature that short and frequent gaps may result in a greater impact on the science for LISA than longer-duration but less 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) as unquantifiable floating points (NaNs), we have shown that ignorantly masking the telemetric data will result in a gap augmentation in the resultant first- and second-generation TDI streams for X , Y , and Z . Since the delay operators are computed using finite-impulse-response filters, we found that the gap augmentation depends linearly 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- and 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 Lagrange-interpolant order of 45 and gap duration ≥ 2.5 s, we observe losses of data lasting approximately 90 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 six in total applied to second generation rather than four 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 longer than 2.5 s. As demonstrated in Fig. 2, we found that a single packet of lost telemetry data (in other words, losing a single sample) may result 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 Table I. We demonstrated that the median effective duty cycle is 84.6% when accounting for both planned and unplanned data gaps, which is approximately 2.6 percentage points higher than what has been quoted in the literature [59]. This is due to the realistic case of overlapping gaps in both the planned and unplanned configurations. Averaging over multiple realizations 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 are given in Table II. We conclude that for planned gaps, short-duration but high-rate data gaps lead to the greatest additional loss of data after performing TDI, compared with the initial data loss in the telemetry, resulting in an effective duty-cycle loss of approximately 0.3%, which corresponds to approximately 4 days 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 approximately 0.01%, i.e., 3 h 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 begin to impact assumptions relating to the process, in particular, whether the between-gap segments are independent or dependent. Spectral leakage, if not fully accounted for in the noise covariance matrix, could also significantly increase the uncertainty in the parameter posterior [19,22,30,31]. No matter the domain—time, frequency, or time-frequency [60,61]—all analyses would be negatively impacted by these short-duration gaps.

A clear take-home message 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 as 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 $\mathbf{X}$; however, other TDI combinations, such as the Sagnac type [42,62,63], might behave differently because they involve shorter delays, and therefore gaps are less spread out. Similarly, techniques such as TDI-Infinity could be better suited to avoiding losses due to gaps [64]. Moreover, the top-right panel of Fig. 2 demonstrates that for kernel orders $\mathcal{K} = 45$, if a single packet of data is lost (of duration approximately 0.25 s), then one could lose a significantly larger amount of data in comparison (say, approximately 86.25 s). Here, we provide a worst-case scenario in which a single missing data value 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.

Nonetheless, 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 that 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 , and 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 gap propagation has not been taken into account here. In the case of long-term clock drifts, the relative frequency offsets between spacecraft clocks can accumulate to significant values over the mission lifetime. For example, a relative offset of $\mathcal{O}(10^{-6})$ will lead to a timing error of $\mathcal{O}(100)$ s over four years (approximately 10^8 s). While such drifts appear as common-mode terms in the applied delays and timestamp 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 [38], or (ii) applying it in a subsequent step after laser-noise reduction [36]. In either case, the required clock measurements are typically reconstructed through sensor fusion of sidebands, carrier tones, and possibly additional references such as pseudorandom noise codes, ground-based signals, or TDI-ranging [65,66]. The propagation of telemetry gaps through this reconstruction is nontrivial. 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, the integration will be forced to restart.

This work forms the foundation for understanding missing data in space-based noise-mitigation pipelines, and we leave further exploration of these concepts for 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 would not have been possible without the use of open-source software—in particular LISAINSTRUMENT, PYTDI, and LISAGLITCH [32,33,56]. We have released our repository on Zenodo [67] and GitHub at https://github.com/OllieBurke/gaps_tdi.

ACKNOWLEDGMENTS

O.B. is deeply thankful to Eleonora Castelli for both the wonderful and insightful conversations regarding LISA simulation and for giving O.B. access to her gap-masking code. O.B. 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.M. is very grateful for the support given by Olaf Hartwig toward 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.M. 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.M. acknowledges Wolfrang Kastau for his hint on using NaNs to propagate and keep track of missing data. O.B. and M.M. are deeply thankful for many insightful discussions with the CU L01 group. In particular, O.B. and M.M. are very grateful to Jean Baptiste-Bayle and to Martin Staab for carefully reviewing the document, which helped to clarify several topics. We also thank the anonymous referees for their very valuable comments, which improved the clarity of the manuscript. M.M. 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.B. and G.W. acknowledge financial support from Grant No. UKRI972 awarded via the UK Space Agency. This work has made use of PYTDI, LISAINSTRUMENT, and LISAGLITCH.

DATA AVAILABILITY

The data that support the findings of this article are openly available [67]; embargo periods may apply.

APPENDIX A: DERIVATION OF GAP AUGMENTATION FOR INTERMEDIARY VARIABLES

1. Preliminaries

In our prescription, missing data will be represented via symbolic NaNs, and 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

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

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

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

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

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

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

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

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 missing telemetry data (gap augmentation) depends on the overall placement of the missing data 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, depending on the order of the interpolant $\mathcal{K}$, the integer time shift N , and the number of contiguous missing data points in the telemetry $N_{\text{NaNs}}^{\text{ifo}}$.

2. 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 Eqs. (A1) and (A7) are *mutually exclusive* and have no common elements. This means that the extra missing data induced via the delay operator are 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

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

with the total number of missing data values 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. We discuss this in more detail in the next section.

3. 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 that the transition point now depends on the choice of integer delay N and the number of missing data samples in the telemetry variables $N_{\text{NaNs}}^{\text{ifo}}$. Similarly, another transition point is the first instance where the missing telemetry dataset is entirely contained within the delayed missing dataset. 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 the 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 the last element in $\mathcal{S}_{\text{NaNs}}^{\text{ifo}}$. A short calculation shows

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

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 this will be the focus of the next section.

4. Regime 3: Complete merging

We now focus on the final (and simplest) case, in which 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

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

and the overall number of missing data points is 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 the pieces together, with the gap augmentations in Eqs. (A11), (A18), and (A19), specific to the constraints in Eqs. (A14) and (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 used follows a near identical procedure to that 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 $N = \lfloor m \times \bar{L} / (c\Delta t) \rfloor$. 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 to those that lead to Eq. (21), we can compute the gap augmentation induced in Eq. (B2):

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

Here, 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 Eqs. (23), (24), and (26), are written in an identical form to (B2). Equation (B3) is used extensively in Sec. III C.

APPENDIX C: MULTIPLE GAPS IN TELEMETRY VARIABLES

For multiple gaps in the telemetry data streams, we can construct the full set of G gaps:

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

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, $n_0^{(i)}$ indicates the first index of the i th gap segment with $i \in \{0, \dots, G-1\}$ for G total data gaps present. Each gap segment $(\mathcal{S}_{\text{NaNs}}^{\text{ifo}})^{(i)}$ contains at most $(N_{\text{NaNs}}^{\text{ifo}})^{(i)}$ points and the segments 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 potentially having 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 are 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 datasets 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 the 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 dataset 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 that in Sec. III C is followed, where one computes the gapped datasets for the $a_{1,2}$, $r_{1,2}$, and $q_{1,2}$ variables defined by Eqs. (23), (24), and (26). The only considerable change is the overall integer delay that must be considered when moving from masked datasets $a_{1,2}$, $r_{1,2}$, and $q_{1,2}$. Due to the extra delays, we must modify the first and last NaN indices of the delayed datasets [Eqs. (C4) and (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 datasets 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 points through propagation of gated telemetry through TDI. We use the formalism presented in this section when we analyze a full-scale four-year long LISA dataset 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 it is applicable to any family of gap types with any duration and rate. Our only requirement is that the missing data values are consistent across all IFO measurements.

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

PYTDI, they can still be quite expensive to compute. Datasets 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 simply to 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 values 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 interpolant $K = 45$, the maximum loss of data is approximately 90 s. This means that if there are multiple (telemetry-based) gaps contained within a 90-s 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 the possibility that we will slightly overestimate the number of missing data elements, which is ultimately conservative at heart.

-
- [1] Monica Colpi, Karsten Danzmann, Martin Hewitson, Kelly Holley-Bockelmann, Philippe Jetzer, Gijs Nelemans, Antoine Petiteau, David Shoemaker, Carlos Sopuerta, Robin Stebbins, *et al.*, LISA definition study report, [arXiv:2402.07571](#).
 - [2] Monica Colpi, Kelly Holley-Bockelmann, Tamara Bogdanovic, Priya Natarajan, Jillian Bellovary, Alberto Sesana, Michael Tremmel, Jeremy Schnittman, Julia Comerford, Enrico Barausse, *et al.*, Astro2020 science white paper: The gravitational wave view of massive black holes, [arXiv:1903.06867](#).
 - [3] Christopher P. L. Berry, Scott A. Hughes, Carlos F. Sopuerta, Alvin J. K. Chua, Anna Heffernan, Kelly Holley-Bockelmann, Deyan P. Mihaylov, M. Coleman Miller, and A. Sesana, The unique potential of extreme mass-ratio inspirals for gravitational-wave astronomy, [arXiv:1903.03686](#).
 - [4] Massimo Tinto and Sanjeev V. Dhurandhar, Time-delay interferometry, *Living Rev. Relativ.* **17**, 6 (2014).
 - [5] J. Aasi, *et al.*, (LIGO Scientific Collaboration), Advanced LIGO, *Class. Quantum Grav.* **32**, 074001 (2015).
 - [6] B. P. Abbott, *et al.*, (LIGO Scientific, and Virgo Collaborations), Observation of gravitational waves from a binary black hole merger, *Phys. Rev. Lett.* **116**, 061102 (2016).
 - [7] Pau Amaro-Seoane, Heather Audley, Stanislav Babak, John Baker, Enrico Barausse, Peter Bender, Emanuele Berti, Pierre Binétruy, Michael Born, Daniele Bortoluzzi, *et al.*, Laser interferometer space antenna, [arXiv:1702.00786](#).
 - [8] Martin Staab, Marc Lilley, Jean-Baptiste Bayle, and Olaf Hartwig, Laser noise residuals in LISA from onboard processing and time-delay interferometry, [arXiv:2306.11774](#).
 - [9] Jan Niklas Reinhardt, Philipp Euringer, Olaf Hartwig, Kohei Yamamoto, and Karsten Danzmann, Time-delay interferometry with onboard optical delays, *Phys. Rev. D* **110**, 082005 (2024).
 - [10] Jean-Baptiste Bayle, Antoine Petiteau, and Hubert Halloin, Effect of filters on the time-delay interferometry residual laser noise for LISA, *Phys. Rev. D* **99**, 084023 (2019).
 - [11] Massimo Tinto, Glenn de Vine, Jeffrey C. Livas, Kirk McKenzie, and Daniel A. Shaddock, Extended time-delay interferometry for LISA, *Phys. Rev. D* **102**, 022003 (2020).
 - [12] Martina Muratore, Time delay interferometry for LISA science and instrument characterization. PhD thesis, Trento U., 2021.
 - [13] Marie-Sophie Hartig, Sönke Schuster, and Gudrun Wanner, Geometric tilt-to-length coupling in precision interferometry: Mechanisms and analytical descriptions, *J. Opt.* **24**, 065601 (2022).
 - [14] S. Paczkowski, R. Giusteri, M. Hewitson, N. Karnesis, E. D. Fitzsimons, G. Wanner, and G. Heinzel, Postprocessing subtraction of tilt-to-length noise in LISA, *Phys. Rev. D* **106**, 042005 (2022).
 - [15] M.-S. Hartig, S. Paczkowski, M. Hewitson, G. Heinzel, and G. Wanner, Postprocessing subtraction of tilt-to-length noise in LISA in the presence of gravitational wave signals, *Phys. Rev. D* **111**, 043048 (2025).
 - [16] Jean-Baptiste Bayle and Olaf Hartwig, Unified model for the LISA measurements and instrument simulations, *Phys. Rev. D* **107**, 083019 (2023).
 - [17] European Space Agency (ESA). Science performance requirements document for LISA, 2022. LISA Science Requirements Document (SciRD).
 - [18] Eleonora Castelli, Quentin Baghi, John G. Baker, Jacob Slutsky, Jérôme Bobin, Nikolaos Karnesis, Antoine Petiteau, Orion Sauter, Peter Wass, and William J. Weber, Extracting gravitational wave signals from LISA data in the presence of artifacts, *Class. Quantum Gravity* **42**, 065018 (2025).
 - [19] Ollie Burke, Sylvain Marsat, Jonathan R. Gair, and Michael L. Katz, Addressing data gaps and assessing noise mismodeling in LISA, *Phys. Rev. D* **111**, 124053 (2025).
 - [20] Ruiting Mao, Jeong Eun Lee, and Matthew C. Edwards, Novel stacked hybrid autoencoder for imputing LISA data gaps, *Phys. Rev. D* **111**, 024067 (2025).
 - [21] Kallol Dey, Nikolaos Karnesis, Alexandre Toubiana, Enrico Barausse, Natalia Korsakova, Quentin Baghi, and Soumen Basak, Effect of data gaps on the detectability and parameter estimation of massive black hole binaries with LISA, *Phys. Rev. D* **104**, 044035 (2021).
 - [22] Quentin Baghi, Ira Thorpe, Jacob Slutsky, John Baker, Tito Dal Canton, Natalia Korsakova, and Nikos Karnesis, Gravitational-wave parameter estimation with gaps in LISA: A Bayesian data augmentation method, *Phys. Rev. D* **100**, 022003 (2019).
 - [23] Jerome Carre and Edward K. Porter, The effect of data gaps on LISA galactic binary parameter estimation, [arXiv:1010.1641](#).
 - [24] Lu Wang, Hong-Yu Chen, Xiangyu Lyu, En-Kun Li, and Yi-Ming Hu, Dealing with data gaps for TianQin with massive black hole binary signal, *Eur. Phys. J. C* **85**, 120 (2025).

- [25] Tyson B. Littenberg and Neil J. Cornish, Prototype global analysis of LISA data with multiple source types, *Phys. Rev. D* **107**, 063004 (2023).
- [26] Michael L. Katz, Nikolaos Karnesis, Natalia Korsakova, Jonathan R. Gair, and Nikolaos Stergioulas, Efficient GPU-accelerated multisource global fit pipeline for LISA data analysis, *Phys. Rev. D* **111**, 024060 (2025).
- [27] Senwen Deng, Stanislav Babak, Maude Le Jeune, Sylvain Marsat, Éric Plagnol, and Andrea Sartirana, Modular global-fit pipeline for LISA data analysis, *Phys. Rev. D* **111**, 103014 (2025).
- [28] Stefan H. Strub, Luigi Ferraioli, Cédric Schmelzbach, Simon C. Stähler, and Domenico Giardini, Global analysis of LISA data with galactic binaries and massive black hole binaries, *Phys. Rev. D* **110**, 024005 (2024).
- [29] P. Whittle, Curve and periodogram smoothing, *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **19**, 38 (1957).
- [30] Colm Talbot, Eric Thrane, Sylvia Biscoveanu, and Rory Smith, Inference with finite time series: Observing the gravitational Universe through windows, *Phys. Rev. Res.* **3**, 043049 (2021).
- [31] Oliver Edy, Andrew Lundgren, and Laura K. Nuttall, Issues of mismodeling gravitational-wave data for parameter estimation, *Phys. Rev. D* **103**, 124061 (2021).
- [32] Jean-Baptiste Bayle, Oskar Hartwig, and Markus Staab, LISA Instrument, Zenodo, 2023.
- [33] M. Staab, J.-B. Bayle, and O. Hartwig, PyTDI (1.2.1). Zenodo (2022), <https://doi.org/10.5281/zenodo.6867012>.
- [34] Marie-Sophie Hartig, Sönke Schuster, and Gudrun Wanner, Geometric tilt-to-length coupling in precision interferometry: Mechanisms and analytical descriptions, *J. Opt.* **24**, 065601 (2022).
- [35] M. Chwalla, K. Danzmann, M. Dovale Álvarez, J. J. Esteban Delgado, G. Fernández Barranco, E. Fitzsimons, O. Gerberding, G. Heinzel, C. J. Killow, M. Lieser, *et al.*, Optical suppression of tilt-to-length coupling in the LISA long-arm interferometer, *Phys. Rev. Appl.* **14**, 014030 (2020).
- [36] Olaf Hartwig and Jean-Baptiste Bayle, Clock-jitter reduction in LISA time-delay interferometry combinations, *Phys. Rev. D* **103**, 123027 (2021).
- [37] Dam Quang Nam, Yves Lemiére, Antoine Petiteau, Jean-Baptiste Bayle, Olaf Hartwig, Joseph Martino, and Martin Staab, TDI noises transfer functions for LISA, *arXiv:2211.02539*.
- [38] Olaf Hartwig, Jean-Baptiste Bayle, Martin Staab, Aurélien Hees, Marc Lilley, and Peter Wolf, Time-delay interferometry without clock synchronization, *Phys. Rev. D* **105**, 122008 (2022).
- [39] Jean-Baptiste Bayle, Olaf Hartwig, and Martin Staab, Adapting time-delay interferometry for LISA data in frequency, *Phys. Rev. D* **104**, 023006 (2021).
- [40] Massimo Tinto, F. B. Estabrook, and J. W. Armstrong, Time-delay interferometry for LISA, *Phys. Rev. D* **65**, 082003 (2002).
- [41] Massimo Tinto and Sanjeev V. Dhurandhar, Time-delay interferometry, *Living Rev. Rel.* **17**, 6 (2014).
- [42] Martina Muratore, Daniele Vetrugno, and Stefano Vitale, Revisitation of time delay interferometry combinations that suppress laser noise in LISA, *Class. Quantum Gravity* **37**, 185019 (2020).
- [43] Michele Vallisneri, Geometric time delay interferometry, *Phys. Rev. D* **72**, 042003 (2005). [Erratum: *Phys. Rev. D* **76**, 109903 (2007)]
- [44] Maude Le Jeune and Stanislav Babak, LISA Data Challenge Sangria (LDC2a) (Version v2) [Data set]. Zenodo (2022), <https://doi.org/10.5281/zenodo.7132178>.
- [45] Martin Staab, Marc Lilley, Jean-Baptiste Bayle, and Olaf Hartwig, Laser noise residuals in LISA from on-board processing and time-delay interferometry, *Phys. Rev. D* **109**, 043040 (2024).
- [46] D. A. Shaddock, B. Ware, R. E. Spero, and M. Vallisneri, Postprocessed time-delay interferometry for LISA, *Phys. Rev. D* **70**, 081101 (2004).
- [47] Martin Staab, Jean-Baptiste Bayle, Olaf Hartwig, Aurélien Hees, Marc Lilley, Graham Woan, and Peter Wolf, Optimal design of interpolation methods for time-delay interferometry, *Class. Quantum Gravity* **42**, 115002 (2025).
- [48] Olaf Hartwig, Instrumental modelling and noise reduction algorithms for the Laser Interferometer Space Antenna. PhD thesis, Leibniz U., Hannover, 2021.
- [49] Markus Otto, Time-delay interferometry simulations for the laser interferometer space antenna. PhD thesis, Leibniz University Hannover, 2015. Doctoral thesis.
- [50] M. Abernathy, A. Amato, A. Ananyeva, S. Angelova, B. Baloukas, R. Bassiri, G. Billingsley, R. Birney, G. Cagnoli, M. Canepa, *et al.*, Exploration of co-sputtered Ta₂O₅–ZrO₂ thin films for gravitational-wave detectors, *Class. Quantum Gravity* **38**, 195021 (2021).
- [51] Olaf Hartwig, Jean-Baptiste Bayle, Martin Staab, Aurélien Hees, Marc Lilley, and Peter Wolf, Time-delay interferometry without clock synchronization, *Phys. Rev. D* **105**, 122008 (2022).
- [52] Waldemar Martens and Eric Joffre, Trajectory design for the ESA LISA mission, *J. Astro. Sci.* **68**, 402 (2021).
- [53] Strictly speaking, missing data of this form are represented by “quiet” NaNs: missing data that do 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.
- [54] Within the gravitational-wave literature, the quantity S is standard for representing the power spectral density of a stochastic process. Here, we choose S to represent a *set* of indices of a discrete random process defining the indices of missing data.
- [55] Martin Staab, Jean-Baptiste Bayle, Olaf Hartwig, Aurélien Hees, Marc Lilley, Graham Woan, and Peter Wolf, Optimal design of interpolation methods for time-delay interferometry, *Class. Quantum Gravity* **42**, 115002 (2025).
- [56] Jean-Baptiste Bayle, Ollie Burke, Eleonora Castelli, and Natalia Korsakova, LISA glitch, 2022.
- [57] Jean-Baptiste Bayle, Marc Lilley, Antoine Petiteau, and Hubert Halloin, Effect of filters on the time-delay interferometry residual laser noise for LISA, *Phys. Rev. D* **99**, 084023 (2019).

- [58] 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.
- [59] Monica Colpi and et al, LISA Definition Study Report, 2024.
- [60] Neil J. Cornish, Time-frequency analysis of gravitational wave data, *Phys. Rev. D* **102**, 124038 (2020).
- [61] Noah Pearson and Neil J. Cornish, Handling data gaps for the next generation of gravitational-wave observatories, [arXiv:2509.05479](https://arxiv.org/abs/2509.05479).
- [62] Martina Muratore, Daniele Vetrugno, Stefano Vitale, and Olaf Hartwig, Time delay interferometry combinations as instrument noise monitors for LISA, *Phys. Rev. D* **105**, 023009 (2022).
- [63] Olaf Hartwig, Jean-Baptiste Bayle, Martin Staab, Aurélien Hees, Marc Lilley, and Peter Wolf, Time-delay interferometry without clock synchronization, *Phys. Rev. D* **105**, 122008 (2022).
- [64] Niklas Houba, Jean-Baptiste Bayle, and Michele Vallisneri, Robust Bayesian inference with gapped LISA data using all-in-one TDI- ∞ , *Class. Quantum Gravity* **42**, 145007 (2025).
- [65] Jean-Baptiste Bayle, Martin Staab, Samuel P. Francis, Emily Rose Rees, Robert Spero, and Gerhard Heinzel, Modulation-assisted time-delay interferometric ranging for LISA, *Class. Quantum Gravity* **42**, 155009 (2025).
- [66] Jan Niklas Reinhardt, Martin Staab, Kohei Yamamoto, Jean-Baptiste Bayle, Aurélien Hees, Olaf Hartwig, Karsten Wiesner, Sweta Shah, and Gerhard Heinzel, Ranging sensor fusion in LISA data processing: Treatment of ambiguities, noise, and onboard delays in LISA ranging observables, *Phys. Rev. D* **109**, 022004 (2024).
- [67] Mar_Mur and Ollie Burke, OllieBurke/gaps_tdi: Impact of gaps on time delay interferometric calculations (1.0.1). Zenodo (2025), <https://doi.org/10.5281/zenodo.17977637>.