

Shape of eccentricity: Rapid classification of eccentric binaries with the wavelet scattering transform

Priscilla Canizares^{*}

Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences,
University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, United Kingdom

Seppe J. Staelens[†]

Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences,
University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, United Kingdom
and Leuven Gravity Institute, KU Leuven, Celestijnenlaan 200D box 2415, 3001 Leuven, Belgium

Isobel Romero-Shaw[‡]

Gravity Exploration Institute, School of Physics and Astronomy, Cardiff University,
Cardiff CF24 3AA, United Kingdom



(Received 9 March 2026; accepted 25 August 2026; published 5 October 2026)

The gravitational wave (GW) detections reported by the LIGO–Virgo–KAGRA (LVK) Collaboration have so far been consistent with quasicircular compact binary coalescences. However, a small fraction of binaries driven to merge through dynamical interactions in dense stellar environments or in field triples may retain measurable orbital eccentricity upon entering the sensitive frequency band of the LVK detectors. A confident measurement of eccentricity in the LVK band would constitute decisive evidence for such dynamically driven mergers. Eccentric waveform models, however, are computationally expensive, so performing production-level inference on *all* detected signals is an inefficient use of resources when eccentric signals are expected to be rare. In this work, we propose a parsimonious strategy that analyzes the *wavelet scattering transform* (WST) coefficients of GW signals as an intermediate step between *detection* and *analysis*: each signal is assessed for the potential presence of eccentricity in order to provide rapid recommendations for which signals should undergo production-level eccentric inference. By compressing the signals and working with lower-dimensional classifiers, this approach reduces computational complexity and thereby speeds up classification. We evaluate the discriminatory power of the WST using a simple one-dimensional convolutional neural network trained on a large set of synthetic waveforms injected into realistic noise. We find that the WST representation enables effective discrimination between eccentric and quasicircular binaries; we discuss the advantages of its compact, multiscale representation and benchmark it against both simpler and more complex classifiers. Using an effective one-body waveform model, our pipeline achieves a detection accuracy of $\sim 64\%$ at a false-alarm rate of 10%, with an AUC of 0.844 and an average precision of 0.876. Finally, we investigate the WST coefficients' ability to distinguish eccentricity from spin-induced precession, and find that it performs well across a range of spin-precession magnitudes. We conclude that the WST is a powerful tool for identifying the fingerprints of eccentricity in compact binary coalescence GW signals.

DOI: [10.1103/fpzz-8mzh](https://doi.org/10.1103/fpzz-8mzh)

I. INTRODUCTION

Gravitational wave (GW) detections made by the LIGO–Virgo–KAGRA (LVK) Collaboration are yielding exciting inferences about the production mechanisms of merging compact binaries, thus providing better constraints on their formation environments. The most recent Gravitational Wave Transient Catalog, GWTC-5 [1], contains 267 candidates with false alarm rate (FAR) $< 1 \text{ yr}^{-1}$ [2]; the properties of this population appear consistent with being produced via a variety of formation mechanisms, including

^{*}Contact author: pc464@cam.ac.uk

[†]Contact author: ss3033@cam.ac.uk

[‡]Contact author: romero-shaw@cardiff.ac.uk

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.

isolated field triples [e.g., Ref. [3]] and dynamical formation in dense star clusters [e.g., Refs. [4,5]]. In contrast to fully isolated binary evolution—which is expected to produce only quasicircular mergers in the LVK frequency range [6]—a small fraction of dynamically-formed and tertiary-influenced mergers should retain measurable eccentricity close to merger [e.g., Refs. [7,8]].

LVK catalogs have not included eccentricity measurements so far: historically, development of quasicircular waveform models was prioritized for detection and analysis of the first detections, leading to inefficient eccentric waveform models. Recently, however, short-author works have used increasingly efficient new eccentric waveform models to place constraints and, in some cases, measure non-negligible eccentricity in compact binary coalescence (CBC) signals [e.g., Refs. [9–15]]. None of the eccentric candidates are confident detections of eccentricity: their interpretation is clouded by missing merger or spin physics in the waveform models, possible confusion with spin-induced precession [16–19], data quality issues [e.g., Ref. [20]], or low signal-to-noise ratio (SNR) [e.g., Ref. [21]]. Nonetheless, with growing evidence for dynamical and/or triple-origin mergers in the data, we could anticipate confident detections of eccentricity in the near future [22]: such formation scenarios have a predicted eccentric merger output of $\sim 10\%$.

There are, however, some hard-to-quantify barriers to these expectations. LVK matched-filter searches are tuned to quasicircular signals, which significantly reduces their sensitivity to eccentric CBC signals [e.g., Refs. [22–24]]. While unmodeled searches may capture some additional high-mass eccentric sources, we remain observationally biased against observing inspiral-dominated eccentric signals [24–26]. Until recently, eccentric waveform models were near-prohibitively expensive, requiring novel methods like importance-sampling with likelihood reweighting [27] or machine learning [10] to make their use in analysis feasible. More efficient eccentric waveform models now exist [e.g., Refs. [28–31]], but they retain the limitations described above, and/or are still much slower than their quasicircular counterparts. Exacerbating this expense, the LVK’s detection efficiency has increased impressively since the first GW detection, and there are now hundreds of new CBC signals per observing run. With eccentric mergers constituting a relatively small fraction of the underlying physical population, and with search methods biased against detecting them, we argue that running eccentric analyses on all LVK candidates may not be well-justified.

It is reasonable, therefore, to propose a computationally efficient selection step in-between detection and eccentric analysis. Machine learning is optimally suited to such a problem, being highly successful in areas requiring pattern recognition and prioritizing speed. Indeed, [31] designed and trained a two-dimensional convolutional neural

network (2D-CNN) to study Q-scans of GW data and identify the presence of eccentricity in the signal. Similarly, [32] proposed a transformer-based approach to detect eccentricity from GW scalograms.

Here, we address this problem using a new approach: employing the wavelet scattering transform (WST) to represent GW strain data. This circumvents the need to process scans or images by operating directly on a compressed representation of the signal. We employ a state-of-the-art eccentric waveform model incorporating higher-order modes [29], and construct an amortized and interpretable WST-based classifier for eccentricity detection.

The WST is a signal-processing technique that builds a multiscale, multiorientation representation of the data in the form of coefficients that capture key structural features [33,34]. The WST has already been demonstrated as a powerful tool to classify and characterise glitches in interferometer strain data, leading to improvements in robust signal detection [35]. Outside of GW data analysis, the WST has been used in cosmology, see Refs. [36–40], as well as metrology, finance and image analysis; see, e.g., Refs. [41–45].

In this work, we find that the WST coefficients show promise for extracting the characteristic features of GW signals from eccentric CBCs in realistic noise. We develop a one-dimensional CNN (1D-CNN) and compare it against classifiers of varying complexity, namely a logistic regression (LR) and a 2D-CNN. While these methods will likely benefit from future improvements—such as altering our reference frequency-based definition of eccentricity to one that defines eccentricity at a fixed number of cycles before merger—we find that they are already able to correctly distinguish between eccentric and quasicircular signals in the majority of the test set, even when the quasicircular signals contain spin-induced precession.

This paper is organized as follows. In Sec. II, we introduce the dataset used in this study and the wavelet scattering transform formulation. We also present our numerical setup: a 1D-CNN classifier for the WST coefficients. In Sec. III, we present our numerical experiments and results. We conclude in Sec. IV by discussing the implications of our findings and outlining future directions.

II. METHODS

A. Waveform generation: SEOBNRv5EHM

We generate synthetic strain data comprising simulated GW waveforms injected into detector noise, using the Bilby library [46,47]. Eccentric waveforms are produced using the SEOBNRv5EHM approximant [29], an effective-one-body, multipolar waveform model describing eccentric binary black hole (BBH) systems with component spins aligned with or against the orbital angular momentum. Priors on the intrinsic parameters are defined within Bilby. The dimensionless component spin magnitudes are drawn

from aligned-spin priors spanning nearly the full physical range, $a \in [0, 0.99]$, and the relativistic anomaly is sampled uniformly in $[0, 2\pi]$. Component masses $m_{1,2}$ are uniformly sampled in $[10, 40]M_\odot$, while the luminosity distance is sampled through the default Bilby prior in the range $[100, 800]$ Mpc—ensuring a distribution that is uniform in comoving volume and source frame time. We initialize the detectors Gaussian noise coloured by O4 noise curves (“realistic noise” hereafter).

The waveform model-defined eccentricity at a reference frequency of 10 Hz, e_{10} , is sampled uniformly within each class: the classification task is to distinguish eccentric systems ($e_{10} \in [0.01, 0.5]$) from noneccentric systems ($e_{10} \in [0.001, 0.01]$). While a log-uniform prior is commonly adopted in GW data analysis, uniform sampling ensures balanced class populations during training. Alternative definitions of eccentricity that do not rely on a reference frequency are also possible, e.g., Refs. [48,49].

Waveforms are generated in the frequency domain over the range 10 Hz–1024 Hz, where the upper cutoff corresponds to half the sampling rate f_s . Each signal has a duration of 128 s and is transformed to the time domain via an inverse fast Fourier transform (FFT). The geocentric merger time is sampled uniformly from 126 ± 0.1 s, and all waveforms are aligned at merger time. Unless otherwise stated, all six multipoles¹ available in SEOBNRv5EHM are included.

The simulated signals are injected into noise realizations of the two Advanced LIGO detectors and the Advanced Virgo detector. An 8 s window around the estimated peak at merger, t_m , is extracted, defined as $[t_m - 7.8 \text{ s}, t_m + 0.2 \text{ s}]$. The resulting strain data segments are whitened, and the network signal-to-noise ratio (NSNR) is computed across the detector network. Samples with $\text{NSNR} < 15$ are discarded to ensure confident detectability while preserving realistic detector noise conditions for training and evaluation.

For this study, we generate approximately 1.2×10^5 waveforms and compute their compressed WST representations. The resulting dataset consists, for each signal and detector, of the scattering coefficients C and the associated time bins T_b , where $T_b \ll T \cdot f_s$ and $T = 8$ s is the total duration of the signal segment considered.

B. Wavelet scattering transform

The WST is a multiresolution signal analysis algorithm [33,34,41,50]: a hierarchical filtering framework that combines wavelet convolutions with modulus and averaging operations to build a multiscale representation of a signal. The resulting scattering coefficients are stable to small deformations, including time warping, and preserve higher-order statistical structure, enabling an informative yet

compact characterization of the data. The coefficients thus encode localized structure and cross-scale dependencies, providing sensitivity to morphological waveform features such as those induced by orbital eccentricity.

In what follows, we present the key aspects of the one-dimensional WST formulation [50] employed in this study. In practice, we implement the transform using the Kymatio package [51], which efficiently computes the required convolutions in the Fourier domain via the FFT.

In one dimension, the WST is parametrized by two quantities, J and Q . Given a sampling frequency f_s , the parameter J sets the maximum (averaging) scale

$$\tau_J = \frac{2^J}{f_s}, \quad (1)$$

that is, the largest time scale over which the scattering coefficients are averaged. Increasing J therefore increases translation invariance while reducing temporal resolution. The parameter Q , on the other hand, specifies the number of wavelets filters per octave and thus determines the sampling density along the frequency axis—larger values of Q provide a finer discretization of scales and hence higher frequency resolution.

The first-order scattering coefficients, S_1 , are obtained by convolving the signal with a wavelet $\psi_{j,q}$, applying the modulus, and a low-pass, ϕ_J , averaging:

$$S_1(t, j_1, q_1) = (|h * \psi_{j_1, q_1}| * \phi_J)(t). \quad (2)$$

These coefficients measure the averaged amplitude of signal fluctuations localized near the scale-frequency index (j_1, q_1) while suppressing high-frequency temporal variations.

Second-order scattering coefficients, S_2 , are computed recursively by applying a second wavelet transform:

$$S_2(t, j_1, q_1, j_2, q_2) = (|h * \psi_{j_1, q_1}| * \psi_{j_2, q_2}| * \phi_J)(t), \quad (3)$$

with the constraint $j_2 > j_1$, to ensure that the second order wavelet probes amplitude modulations occurring at longer timescales than those captured by the first wavelet. Hence, S_2 coefficients recover information lost during the first averaging step and encode correlations between structures at different scales. Averaging with ϕ_J ensures that all scattering coefficients are invariant to translations up to a scale τ_J . We refer the reader to Ref. [50] and Appendix B for a more detailed explanation on the one-dimensional wavelet scattering transform.

In a similar way, higher-order coefficients can be defined recursively. However, their amplitudes decrease rapidly with order, and most of the relevant information is captured by the first two orders [33]. Accordingly, in this work we restrict our analysis to orders ≤ 2 . Figure 1 shows that the WST representation captures higher-order structure and

¹These correspond to $(\ell, m) = (2, 2), (2, 1), (3, 2), (3, 3), (4, 3)$, and $(4, 4)$.

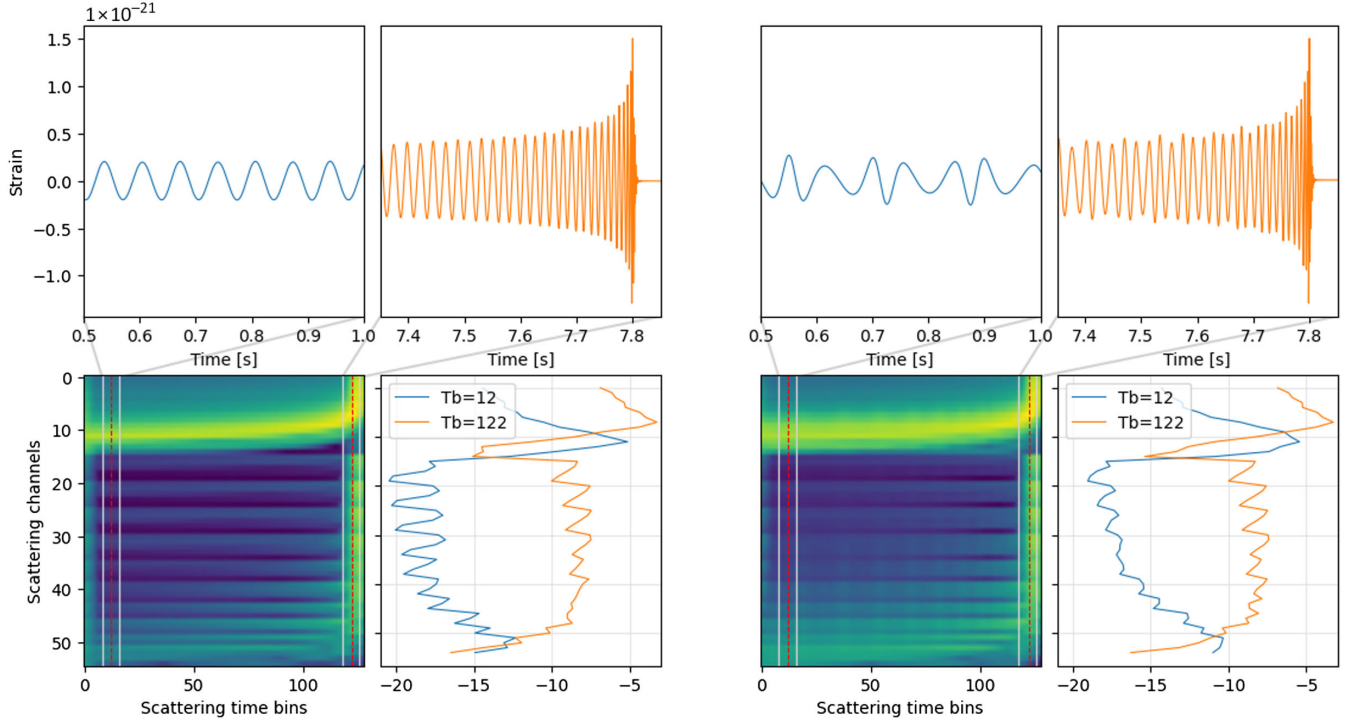


FIG. 1. Illustrative example of a (left 4 panels) quasicircular and (right 4 panels) eccentric ($e_{10} = 0.3$) waveform, with zero noise, and their corresponding $(J, Q) = (7, 2)$ WST coefficients (on a logarithmic scale). The binary masses are 12 and $18M_{\odot}$, at a luminosity distance of 100 Mpc. The top row shows two 0.5 s windows of the waveform, where the first represents the inspiral and the second the merger. The bottom row represents all WST coefficients in the left panel, as well as a slice at two time bins T_b in the right panel, corresponding to the different phases in the top row. From these plots, it is clear that, as expected, the WST coefficients representing the eccentric binary distribute the power across the channels more evenly; i.e., many of the WST coefficients have increased values compared to the circular case. It should be noted that the top 15 channels in the bottom row represent first-order scattering coefficients, and the rest second-order.

amplitude modulations, including those induced by orbital eccentricity.

C. Numerical setup: 1D-CNN

CBCs with nonzero orbital eccentricity produce GW signals whose time-frequency structure differs from that of quasicircular inspirals [48,52,53]. Eccentric systems radiate power into multiple harmonics of the orbital frequency, producing a richer multiresolution structure that the WST is able to capture [54]. Figure 1 illustrates this by comparing the WST coefficients of a circular and an eccentric system.

In what follows, we present a one-dimensional convolutional neural network (1D-CNN) classifier designed to analyze the combined WST coefficients from the individual GW detectors and to exploit their multiresolution structure for eccentricity classification. Our main goal is to develop a parsimonious strategy that reduces computational complexity and thereby speeds up classification: by compressing the signals and working with lower-dimensional classifiers, we avoid the need for more costly intermediate steps such as constructing spectrograms and training complex CNN architectures. We benchmark our approach against two other classifiers of contrasting complexity: a linear model

(LR) and a nonlinear model (2D-CNN), an architecture originally designed for image analysis, which we discuss in Appendix C.

1. One-dimensional CNN

A key consideration is that each detector in the network has distinct characteristics, most notably different noise power spectral densities and antenna response patterns. To account for these differences while still learning common signal features, the architecture, Fig. 2, takes the WST representation of the data from each detector as input and applies a shared convolutional block to their WST features, ensuring that the learned filters capture detector-independent temporal signatures. The resulting detector embeddings are subsequently fused into a joint representation that captures the common source information encoded in their data. Concretely, the input to the 1D-CNN is a WST tensor of shape $X_n \in \mathbb{R}^{B \times D \times C \times T_b}$, where B is the data batch size, D the number of detectors, C the number of WST frequency channels, and T_b the number of time bins. For each detector d , the slice $X_{n,d} \in \mathbb{R}^{B \times C \times T_b}$ is processed independently.

To make the time axis consistent across events on the same detector, the synthetic data are aligned at merger.

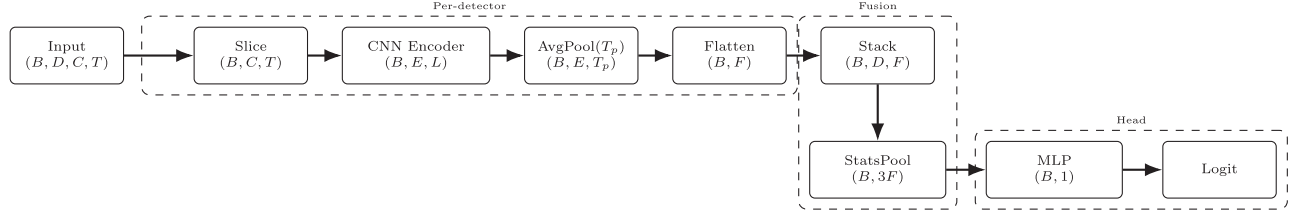


FIG. 2. Block diagram of the 1D-CNN designed for eccentricity detection. Data from different detectors are encoded separately by a shared one-dimensional convolutional layer with three blocks and intermediate max-pooling for temporal downsampling. Feature maps are averaged into T_p bins, producing fixed-length detector embeddings. These are fused (joined) using first-order statistics, and the resulting representation is classified by a shallow MLP to yield the final probability value.

This alignment ensures that corresponding temporal features occur at the same relative positions in the scattering representation, allowing the network to learn consistent temporal patterns associated with eccentricity. Although times of arrival are slightly offset between detectors for the same event, as is also the case for real detections, this effect is mitigated by the modulus operation of Eqs. (2) and (3).

The network maps the WST data into a lower-dimensional representation for each event and detector across channels and time bins. It consists of three convolutional blocks, each composed of a 1D-convolutional layer followed by batch normalization and a *ReLU* activation, with *max-pooling* applied after the first two blocks to reduce the temporal dimension. This encoder structure progressively reduces the number of feature channels: $C \rightarrow 64 \rightarrow 32 \rightarrow 16$, using kernel sizes (7, 5, 3). The resulting encoding is

$$\mathbf{H}_{n,d} \doteq f_{\text{cnn}}(\mathbf{X}_{n,d}) \in \mathbb{R}^{E \times L}, \quad (4)$$

where $E = 16$ is the number of embedding channels and $L = \text{floor}(T_b/2)$ denotes the reduced temporal dimension. To obtain a fixed-length, merger-aligned summary, we apply adaptive average pooling to T_p time bins,

$$\tilde{\mathbf{H}}_{n,d} = \text{Pool}_{T_p}(\mathbf{H}_{n,d}) \in \mathbb{R}^{E \times T_p}, \quad (5)$$

effectively dividing the time axis into T_p coarse temporal segments and producing a compact representation of the scattering coefficients. The pooled representation is then flattened to yield the final detector embedding

$$\mathbf{e}_{n,d} = \text{vec}(\tilde{\mathbf{H}}_{n,d}) \in \mathbb{R}^{E \cdot T_p}. \quad (6)$$

This feature representation captures learned temporal correlations across WST channels at multiple time scales and across detectors.

To construct a joint detector representation, we stack the individual detector embeddings into a detector-set representation

$$\mathbf{Z}_n = [\mathbf{e}_{n,1}, \dots, \mathbf{e}_{n,D}] \in \mathbb{R}^{D \times (E \cdot T_p)}, \quad (7)$$

and pool across the detector dimension to form the final feature vector $\mathbf{p}_n \in \mathbb{R}^{3(E \cdot T_p)}$. This vector is then passed to a two-layer multilayer perceptron (with *ReLU* activation and dropout of 0.3), which outputs a scalar, s , defining the eccentric versus noneccentric classification as explained in Sec. II C 2. With default parameters— $T_p = 6$, an intermediate multilayer perceptron (MLP) of size 128, and $(J, Q) = (7, 2)$ —the model has 74 321 trainable parameters.

2. Cross-validation, threshold calibration, and evaluation

To evaluate the performance of the 1D-CNN approach, we reserve a stratified hold-out test set comprising 20% of the data, which is excluded from both training and threshold selection. This ensures that neither the model parameters nor the operating thresholds are influenced by the test data. The remaining 80% is used for stratified K -fold cross-validation with $K = 10$, reducing sensitivity to the choice of validation split.

The network is trained as binary classifier, eccentric ($e_{10} \in [0.01, 0.5]$) versus noneccentric ($e_{10} \in [0.001, 0.01]$), by minimizing the binary cross-entropy loss. The threshold value $e_{10}^{\text{thr}} = 0.01$ adopted for “circular” events is a practical choice. Optimization is performed using the Adam optimizer with a learning rate of 5×10^{-4} and a batch size of 1000. To mitigate overfitting and promote generalization, early stopping is applied independently within each fold based on the validation AUC (area under the receiver operating characteristic curve), with a patience of five epochs.

Since signals confidently identified as eccentric are expected to constitute only a small fraction of detected CBCs, we evaluate the model at fixed operating points along the FAR axis—also known in the machine-learning literature as the false positive rate (FPR). Fixing the FPR enables a fair comparison across classifiers, since the same score threshold can correspond to very different classification rates for different classifiers. We define the FPR as the fraction of negative samples assigned to the positive class,

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} = \Pr(\hat{y} = 1 | y = 0). \quad (8)$$

Note that we use FPR in the dimensionless classification sense (the fraction of negatives misclassified as positives), not the per-unit-time event rate conventional in LVK searches. At a fixed FPR, we quantify performance by the fraction of positive samples correctly recovered, the true positive rate (TPR),

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \Pr(\hat{y} = 1 | y = 1). \quad (9)$$

For each fold, the decision threshold s^* is calibrated on the validation set against the distribution of negative scores: to achieve a target FPR of α , we set the threshold s^* to the $(1 - \alpha)$ quantile of the negative validation scores. For example, for a target FPR of 1%, s^* is chosen so that 99% of the negative validation scores fall below it. The calibrated thresholds are then applied to the held-out test set to measure the realized FPR and TPR; any deviation of the realized FPR from its target reflects how well the threshold generalizes from validation to test data. Finally, to quantify overall discriminative performance independent of any specific operating point, we additionally report the AUC and the average precision (AP). In Appendix C we report the same metrics for the LR and 2D-CNN classifiers for comparison.

III. NUMERICAL EXPERIMENTS AND RESULTS

In this section, we present numerical experiments evaluating eccentricity detection performance using the WST of the simulated GW strain data described in Sec. II A. The primary results are obtained with the 1D-CNN (Sec. II C 1). For completeness and comparison, we also report results obtained with a 2D-CNN classifier and LR in Appendix C. Across our experiments, the 1D-CNN architecture applied to WST coefficients with $(J, Q) = (7, 2)$ emerges as the best-performing configuration overall, and we adopt it as our reference model in what follows. We note that a suitable choice of J depends strongly on the length of the data considered and the sampling rate; see Appendix B for a detailed explanation. Metrics for this configuration are reported in Table I, for $\text{FPR} = 10^{-1}, 10^{-2}, 10^{-3}$ —alongside results for $(J, Q) = (7, 1)$, which illustrate the benefit of increasing the scattering depth Q to incorporate additional frequency-dependent structure. Further decreasing J or increasing Q can yield modest improvements, but at the cost of a rapidly growing WST representation (see Table II reporting best-fold metrics for each model); the choice $(J, Q) = (7, 2)$ shows the best trade-off between performance and feature dimensionality. Further validation metrics are shown in Appendix C.

A. Characterizing eccentric systems

Weakly eccentric systems—i.e. binaries with eccentricities close to the class threshold $e_{10}^{\text{thr}} = 0.01$ —are

TABLE I. Comparison of detection performance for WST data computed on the same dataset, but with $Q = 1, 2$ and fixed $J = 7$. Shown are the mean and standard deviation across $K = 10$ folds for $\text{FPR} = \{0.1, 0.01, 0.001\}$. Results are obtained with the 1D-CNN model. The relative difference is given by $\Delta = ((M_{Q=2} - M_{Q=1})/M_{Q=1}) \times 100$ computed using the column means.

	$Q = 2$	$Q = 1$	$\Delta\%$
Validation AUC	0.829 ± 0.010	0.823 ± 0.004	+0.73
Validation AP	0.862 ± 0.010	0.856 ± 0.004	+0.70
Test AUC	0.831 ± 0.011	0.824 ± 0.003	+0.85
Test AP	0.863 ± 0.010	0.856 ± 0.003	+0.82
Validation TPR			
FPR 0.1	0.599 ± 0.023	0.580 ± 0.011	+3.28
FPR 0.01	0.422 ± 0.026	0.408 ± 0.014	+3.43
FPR 0.001	0.333 ± 0.033	0.326 ± 0.025	+2.15
Test TPR			
FPR 0.1	0.600 ± 0.028	0.585 ± 0.010	+2.56
FPR 0.01	0.424 ± 0.026	0.411 ± 0.012	+3.16
FPR 0.001	0.337 ± 0.032	0.331 ± 0.024	+1.81

expected to be the most difficult to distinguish from quasicircular systems, and therefore the most likely to be misclassified. To quantify this, we focus on the eccentric binaries in the test set ($e_{10} > e_{10}^{\text{thr}}$) and extract the subset of true positives (TPs), i.e. the eccentric binaries correctly classified as eccentric at a target FPR of 10%. These TPs are the eccentric binaries whose classification score $s > s^*$ calibrated to this FPR (Sec. II C 2). The left panel of Fig. 3 shows that the eccentricity distribution of the TP subset is skewed toward higher values of e_{10} , consistent with the expectation that weakly eccentric systems are harder to distinguish from circular ones. The classifier thus identifies highly eccentric binaries with greater confidence, with misclassifications concentrated near the class boundary.

The strength of a GW signal depends primarily on the binary masses and luminosity distance, which together set the SNR and, in turn, the classifier performance. To investigate how these physical parameters influence eccentricity classification, we examine the dependence on the total mass M , the mass ratio $q = m_2/m_1 \leq 1$, and the NSNR.

TABLE II. Performance of the 1D-CNN classifier for different choices of the WST parameters J and Q in the data. Performance is reflected by the threshold-agnostic metrics AUC and AP on the test set, using the best fold. Finally, the TPR is given for a target FPR of 10%. Bold marks the reference configuration, $(J, Q) = (7, 2)$, adopted in this work.

(J, Q)	AUC	AP	$\text{TPR}_{0.1}$
(7, 2)	0.844	0.876	0.636
(8, 2)	0.832	0.863	0.597
(6, 2)	0.845	0.876	0.622

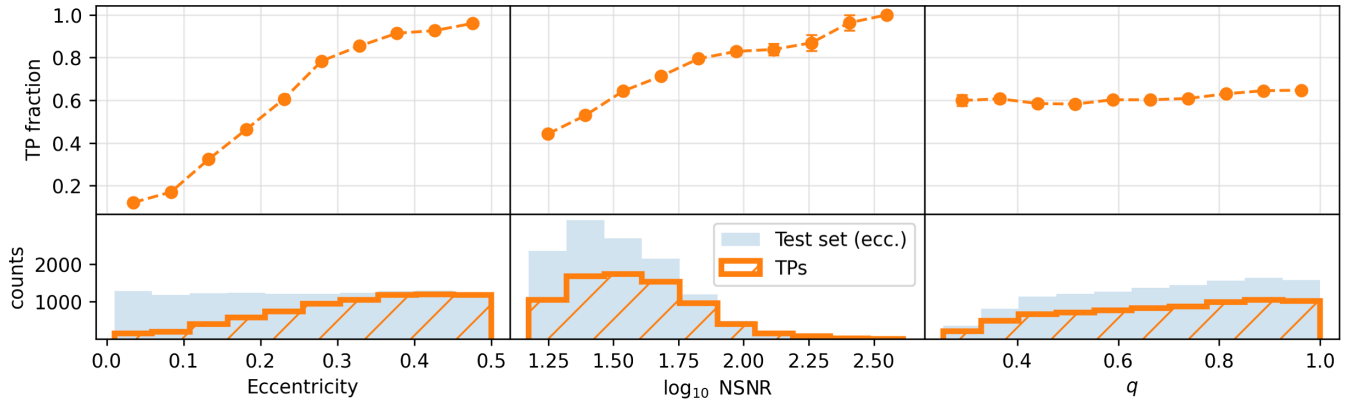


FIG. 3. Performance of the reference 1D-CNN classifier, with the decision threshold s^* calibrated to a FPR of 10% (Sec. II C 2), on the eccentric binaries in the test set. The performance is shown for different bins of the (left) eccentricity e , (middle) network SNR and (right) mass ratio q . The top row shows the fraction of TPs in each bin, the bottom row shows a histogram of the test set and the TPs.

Figure 3 shows the distribution of TPs as a function of NSNR and q : the mass ratio has only a minor effect on performance, whereas the NSNR is highly influential, with low-SNR events exhibiting substantially lower TP rates. The influence of the total mass is discussed further in Sec. IV A. In addition, we characterize classifier performance through the minimum detectable eccentricity $e_{10}^{\min}$. For a given NSNR, we define it as the smallest eccentricity at which the TPR reaches a target detection probability P_{det} :

$$e_{10}^{\min}(\text{NSNR}) := \inf\{e | \text{TPR}(e, \text{NSNR}) \geq P_{\text{det}}\}. \quad (10)$$

Estimating $e_{10}^{\min}$ requires a smooth model of the detection probability across the (e, NSNR) plane. To this end, we apply the `GaussianProcessClassifier` (GPC) from `SCIKIT-LEARN` [55]: as input, we randomly sample 2500 eccentric binaries from the test set² and use the corresponding values of e and $\log_{10}(\text{NSNR})$. The labels encode the detection outcome: a value of 1 indicates that an eccentric binary was correctly identified by the 1D-CNN classifier at the fixed FPR (i.e., its score s exceeds s^* ; see Sec. II C 2), and 0 otherwise. The GPC fits a continuous latent function $f(e, \log_{10}(\text{NSNR}))$ and maps it to a detection probability through a logistic sigmoid. We adopt a composite kernel—a constant kernel multiplied by a radial basis function kernel with an independent length-scale for each input—and optimize its hyperparameters by maximizing the marginal log-likelihood during fitting. The minimum detectable eccentricity at a fixed detection probability P_{det} is then given by the corresponding isocontour of the fitted detection-probability surface. As shown in Fig. 4, $e_{10}^{\min}$ decreases monotonically with NSNR, as expected, for both $P_{\text{det}} = 0.9$ and $P_{\text{det}} = 0.5$.

²We do not use the full test set due to computational overhead. We checked that increasing the number of samples does not significantly alter the results.

B. Contribution of higher multipoles

To investigate the impact of higher multipoles on eccentricity detection, we train the 1D-CNN classifier on `SEOBNRv5EHM` waveforms restricted to the dominant quadrupole mode, $(\ell, m) = (2, 2)$, and evaluate its performance on data that include all multipoles. The results are summarized in Table III for different FPRs, $10^{-1}, 10^{-2}, 10^{-3}$. We can see that with training on quadrupole-only waveforms the WST representation yields robust classification performance, with test AUC ≈ 0.83 and stable detection efficiency (TPR) across all operating points. This indicates that the quadrupole mode captures the dominant eccentricity-dependent structure present in the scattering coefficients. We find that the quadrupole-only and all-modes models exhibit statistically consistent performance

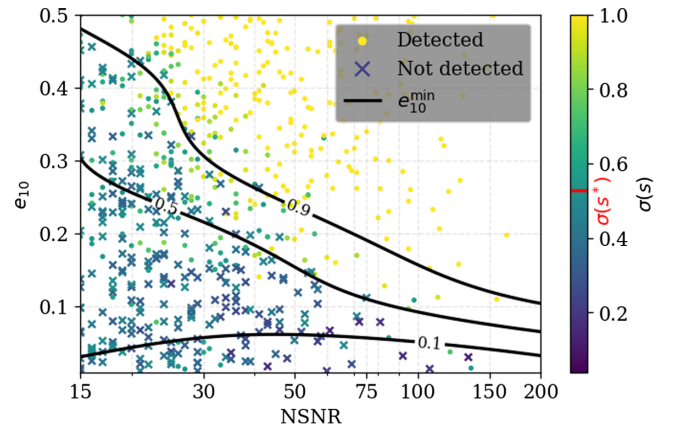


FIG. 4. Estimate of the minimum detectable eccentricity $e_{10}^{\min}$, Eq. (10), at detection probabilities $P_{\text{det}} = 0.9, 0.5, 0.1$ as a function of NSNR, for a fixed FPR = 10%. The color bar corresponds to the classifier output s , rescaled to $[0, 1]$ with a sigmoid σ , and s^* is the decision threshold for this FPR (Sec. II C 2). The points correspond to a sample of the data: dots (crosses) indicate that the binary was (in)correctly classified as eccentric (circular) by the reference 1D-CNN.

TABLE III. Performance comparison between the quadrupole-only and full (all modes) 1D-CNN model, for $\text{FPR} = \{0.1, 0.01, 0.001\}$. Results are reported over $K = 10$ folds, as mean $\pm$ standard deviation. The relative difference is defined as $\Delta = [(M_{\text{quad}} - M_{\text{all}})/M_{\text{all}}] \times 100$, computed from the column means.

	Quadrupole	All modes	$\Delta(\%)$
Validation AP	0.864 ± 0.014	0.862 ± 0.010	+0.23
Test AUC	0.828 ± 0.013	0.831 ± 0.011	-0.36
Test AP	0.861 ± 0.012	0.863 ± 0.010	-0.23
Validation TPR			
FPR 0.1	0.61 ± 0.03	0.60 ± 0.02	+1.67
FPR 0.01	0.43 ± 0.04	0.42 ± 0.03	+2.38
FPR 0.001	0.34 ± 0.04	0.33 ± 0.03	+3.03
Test TPR			
FPR 0.1	0.60 ± 0.03	0.60 ± 0.03	+0.00
FPR 0.01	0.43 ± 0.04	0.42 ± 0.03	+2.38
FPR 0.001	0.34 ± 0.04	0.34 ± 0.03	+0.00

across all metrics. The all-modes model achieves marginally lower test AUC and AP, with relative differences of 0.36% and 0.23%, respectively, but these are small compared to the fold-to-fold variation and do not indicate a statistically significant improvement.³ The true-positive rates at fixed FPR are similarly consistent. On the validation set, the quadrupole-only model yields slightly higher TPR at all FPR thresholds, while on the independent test set (which includes all modes) the two models agree to within $\lesssim 1.45\%$ across all operating points.

The ability of the 1D-CNN to achieve strong performance using only quadrupole information reflects its capacity to extract nonlinear temporal features from the scattering coefficients. At the same time, the absence of improvement from including higher multipoles suggests that their contribution to eccentricity detection is subdominant in the mass range $10 - 40M_{\odot}$.

C. Binaries with spin-induced precession

Eccentricity and spin-induced precession can produce qualitatively similar amplitude and phase modulations in GW signals, [e.g., Ref. [16]], so a classifier trained to identify eccentricity may instead respond to precession-induced features. To assess this potential degeneracy, we apply our classifier to a test set of quasicircular, precessing BBHs generated with the SEOBNRv5PHM approximant [57]. Both the SEOBNRv5PHM (noneccentric, precessing-spin) and SEOBNRv5EHM (eccentric, aligned-spin) models share the same aligned-spin, quasicircular SEOBNRv5HM baseline used during training. This common baseline limits modelling systematics differences between training and test data; in particular, the mismatch between SEOBNRv5PHM

and SEOBNRv5HM is Ref. [58] $\sim 10^{-6}$. We adopt the standard BBH priors defined in Bilby, with the reference frequency for spin definitions fixed at 10 Hz.

We quantify the degree of spin precession using the effective precession parameter χ_p [57],

$$\chi_p = \frac{1}{B_1 m_1^2} \max(B_1 m_1^2 \chi_{1,\perp}, B_2 m_2^2 \chi_{2,\perp}), \quad (11)$$

where $m_1 \geq m_2$ are the component masses, $B_i = 2 + 3m_{\neq i}/m_i$, and $\chi_{i,\perp}$ is the component of the dimensionless spin perpendicular to the orbital angular momentum. To ensure broad coverage of the precession parameter space, we draw spin magnitudes from β -distributions with parameters $\alpha = \beta = 0.5$, which preferentially sample both small and near-extremal spins and thereby enhance coverage in χ_p .

To understand how precession affects the scattering representation, we examine the first-order scattering coefficients S_1 , for precessing waveforms. From Eqs. (2), (A7), and (B4) follows that S_1 is bounded by

$$S_1(\lambda, t) \leq C \int dt \int dv \left| \sum_{m'} d_{mm'}^e(\beta) A_{\ell m'}(t) e^{-\frac{(t-v)^2}{2\sigma^2}} \right|, \quad (12)$$

where C is a constant. This makes explicit how the mode-mixing coefficients $d_{mm'}^e(\beta)$ are captured by the scattering representation. For precessing systems, the precession angle $\beta(t)$ evolves on the precession timescale and generates the additional amplitude and phase modulations characteristic of precessing systems. In the aligned-spin limit relevant to SEOBNRv5EHM, by contrast, β is constant and these time-dependent modulations are absent, yielding qualitatively different WST representations in the two cases (see Appendix A).

Figure 5 shows the fraction of precessing binaries misclassified as eccentric by the reference 1D-CNN classifier, calibrated to a FPR of 10%. Across all values of χ_p , the FP fraction remains $\lesssim 10\%$, consistent with the targeted FPR. We note that this fraction does not increase significantly with χ_p , even though precessing systems are absent from the training data. The classifier therefore does not strongly confuse precession-induced modulations with eccentricity signatures⁴: its false-positive rate stays consistent with the calibrated operating point, indicating that the WST-based representation captures features that distinguish eccentricity from spin-induced precession.⁵

⁴Similar results were reported by Ref. [19] after submission of this manuscript.

⁵Repeating the analysis with a target FPR of 1% yields the same conclusion: the FP fraction on the precessing dataset remains consistent with the target FPR.

³These findings are consistent with [56] published after our work was submitted.

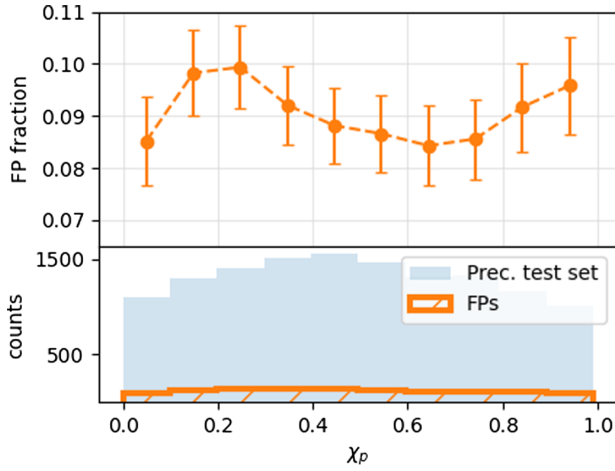


FIG. 5. Bottom: histogram of χ_p values in the precessing, quasicircular binary dataset, together with the subset classified as false positives (FPs) by the reference 1D-CNN tuned to a FPR of 10%. Top: fraction of FPs in each bin of the bottom panel.

IV. DISCUSSION AND CONCLUSIONS

A. Definition of eccentricity

Eccentricity decreases during the inspiral due to GW emission [6], so its observational signatures are most prominent at earlier times. Lower-mass binaries remain in the detector band for longer before merger, and one might therefore expect eccentricity to be easier to detect in such systems.

Our results initially suggest the opposite: classifier performance appears to improve for higher-mass binaries (Fig. 6). This is explained by a correlation between total mass and NSNR in our dataset, since higher-mass binaries tend to produce stronger signals, in part because of the restricted luminosity-distance range. To disentangle these effects, we examine the TP fraction as a function of total mass while restricting to lower-SNR events ($\text{NSNR} < 30$), reducing this correlation. Under this restriction, higher-mass binaries become more difficult to classify correctly, as shown in Fig. 6. This is consistent with expectations, since higher-mass binaries have shorter observable inspirals and therefore less visible eccentricity signatures.

This counterintuitive behavior as a function of total mass may be exacerbated by the definition of eccentricity used to label the dataset, namely the eccentricity evaluated at a fixed reference frequency of 10 Hz. We capture only the last 8 s of the signal in the WST, which, for lower-mass binaries, omits much of the early inspiral. This both reduces the SNR within the analyzed portion of the signal and shifts the analysis to a less eccentric part of the waveform, since eccentricity decreases as the orbital frequency increases. As a result, binaries labelled eccentric at 10 Hz may appear nearly circular over the portion of the signal we pass to the classifier, increasing the likelihood of misclassification. We will mitigate this effect by extending to longer segments in future work, discussed further below.

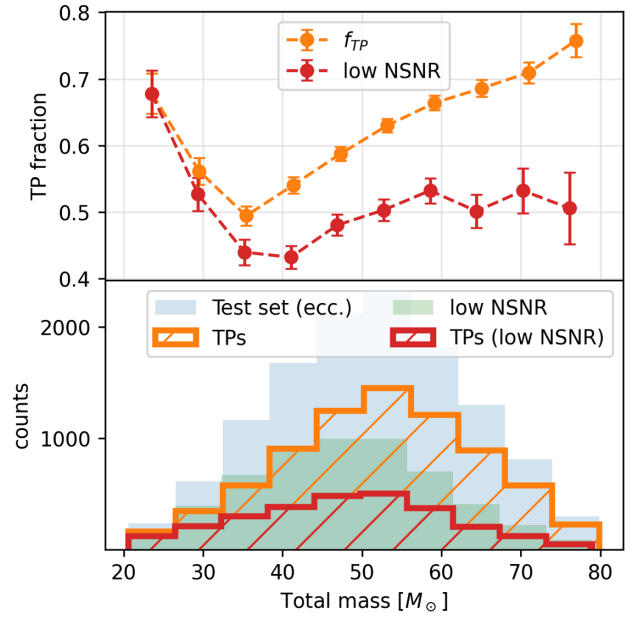


FIG. 6. (Bottom) Distribution of total mass within the test set of eccentric binaries, and within the subset with low NSNR (< 30). Hatched histograms show the true positives (TPs) in each. (Top) Fraction of TPs in each bin of the bottom panel, for the full eccentric set and the low-NSNR subset.

B. Conclusions and future work

In this work, we have demonstrated that the WST provides a compressed, robust, and physically meaningful representation for identifying orbital eccentricity in gravitational-wave signals embedded in realistic detector noise. To exploit these properties, we designed an interpretable and parsimonious 1D-CNN classifier that achieves strong detection performance while remaining computationally efficient, suggesting that the essential eccentricity information is already captured by the wavelet scattering coefficients.

Among the physical factors that influence classifier performance, the signal-to-noise ratio is the dominant driver of the TPR, while the total mass and mass ratio play secondary roles once their correlations with signal strength are controlled for.

We further showed that strong performance can be achieved using quadrupole information alone, indicating that the dominant eccentricity signatures are encoded in the nonlinear features of the scattering coefficients. The lack of improvement from including higher multipoles suggests that their contribution is subdominant in the mass range $10 - 40 M_\odot$, although they are expected to become more important for systems with larger mass ratios or stronger precession-eccentricity correlations.

We have also investigated whether our approach can distinguish eccentricity from spin-induced precession, applying it to a test set of quasicircular waveforms generated with SEOBNRv5PHM [57], a state-of-the-art model that allows for misaligned spins. We find that classifier

performance is consistent across the full range of χ_p , indicating that the WST is an informative representation for separating eccentric from precessing binaries.

This work can be extended in several directions. The proposed framework could be generalized to perform joint detection and parameter estimation and to cover a broader mass range. Alternative eccentricity definitions—better aligned with the observable signal—could also be explored, together with an assessment of performance over wider regions of parameter space.

ACKNOWLEDGMENTS

The authors thank Adhrit Ravichandran and Dr. Rudy Morel for their helpful comments, and especially Dr. Christopher Moore for his valuable insights. We are also grateful to Prof. Carola-Bibiane Schönlieb for her support, and to the Bristol Supercomputing Centre for computational resources. Additional computations were done on the CSD3 cluster (Cambridge). The authors also thank the referees for their thorough and insightful feedback on our manuscript. P.C. is supported by a Daphne Jackson Fellowship sponsored by the Alan Turing Institute. S.J.S. is supported by the Centre for Doctoral Training at the University of Cambridge, funded by STFC. I.R.-S. acknowledges support from the STFC Ernest Rutherford Fellowship, Grant No. UKRI2423.

DATA AVAILABILITY

The code that supports the findings of this article, together with example data, is openly available in the WST-Eccentricity repository [59,60]. The full synthetic datasets (the simulated waveforms and their wavelet-scattering-transform coefficients) cannot be made publicly available because of their large volume; they can be regenerated with the released code and publicly available waveform-generation software, and are available upon reasonable request from the authors.

APPENDIX A: SEOBNRV5EHM AND SEOBNRV5PHM

A gravitational-wave waveform can be decomposed as [61,62]

$$h(t; \iota, \varphi_0) = \frac{1}{D_L} \sum_{\ell, m} h_{\ell m}(t) {}_{-2}Y_{\ell m}(\iota, \varphi_0), \quad (\text{A1})$$

where D_L is the luminosity distance and (ι, φ_0) describe the source orientation.

In the effective-one-body (EOB) formalism, the $h_{\ell m}$ modes are constructed from the orbital dynamics [63–66]. SEOBNRV5EHM [29] and SEOBNRV5PHM [57] have the same EOB backbone dynamics but differ in their treatment of eccentricity and spin precession respectively.

SEOBNRV5EHM is designed to incorporate eccentricity effects, and in the circular limit its modes take a simple form [29,65,67]:

$$h_{\ell m}^{\text{EHM}}(t) = A_{\ell m}(t) e^{-i\phi_{\ell m}(t)}, \quad (\text{A2})$$

with phase

$$\phi_{\ell m}(t) = m\Phi_{\text{orb}}(t) + \delta_{\ell m}(t), \quad (\text{A3})$$

where $\Phi_{\text{orb}}(t)$ is the orbital phase and $\delta_{\ell m}(t)$ contains post-Newtonian and non-quasicircular corrections [68,69]. Each mode is therefore governed by a single dynamical phase variable,

$$h_{\ell m}^{\text{EHM}}(t) = A_{\ell m}(t) e^{-i[m\Phi_{\text{orb}}(t) + \delta_{\ell m}(t)]}. \quad (\text{A4})$$

On the other hand, SEOBNRV5PHM incorporates precession by constructing the waveform in a co-precessing frame and rotating to the inertial frame [57,66]:

$$h_{\ell m}^{\text{PHM}}(t) = \sum_{m'} D_{mm'}^{\ell}(\alpha, \beta, \gamma) h_{\ell m'}^{\text{copr}}(t), \quad (\text{A5})$$

where $D_{mm'}^{\ell}$ are Wigner rotation matrices parametrized by the time-dependent Euler angles (α, β, γ) [70]. The effective spin precession, χ_p , Eq. (11), measures the magnitude of the spin components perpendicular to the instantaneous orbital angular momentum $\hat{\mathbf{L}}(t)$ that drive precession, and therefore sets the time scale of the dependence of the Euler angles (α, β, γ) appearing in Eq. (A5), which in turn control the strength of mode mixing and the resulting amplitude and phase modulations in the inertial-frame waveform modes [57,66,71,72].

The co-precessing modes retain the aligned-spin structure,

$$h_{\ell m}^{\text{copr}}(t) = A_{\ell m}^{\text{copr}}(t) e^{-i[m\Phi_{\text{copr}}(t) + \delta_{\ell m}(t)]}. \quad (\text{A6})$$

Substituting yields the inertial-frame modes,

$$h_{\ell m}^{\text{PHM}}(t) = \sum_{m'} D_{mm'}^{\ell}(\beta) A_{\ell m'}^{\text{copr}}(t) e^{-i[m\alpha + m'\gamma + m'\Phi_{\text{copr}} + \delta_{\ell m'}]}. \quad (\text{A7})$$

We can see that while SEOBNRV5EHM contains independent modes characterized by a single orbital phase, the precession effects in the SEOBNRV5PHM modes contain additional phases (α, γ) and mixing controlled by β that produce additional amplitude and phase modulations.

APPENDIX B: THE 1D WAVELET SCATTERING TRANSFORM

Following the notation used in Kymatio [50,51], the scattering transform is defined relative to the duration of the input signal. In our case, the WST is applied to a discretized detector strain signal $h(t)$ of duration T and sampling frequency f_s , so that the total number of time samples is $N = Tf_s$. The Fourier frequency resolution of such a segment is set by the base frequency

$$\nu_0 = \frac{1}{T}, \quad (\text{B1})$$

i.e., the spacing between adjacent Fourier bins. We stress that ν_0 is the resolution of the segment, not the lowest wavelet frequency: as shown below, the latter is instead controlled by the maximum scale J .

The scattering filter bank is composed of wavelets indexed by $\lambda = (j, q)$, where $j = 1, \dots, J$ labels the octave (dyadic scale 2^j) and $q = 0, \dots, Q - 1$ labels the position within each octave. Following the Kymatio [51] convention, the bank is anchored at the high-frequency (Nyquist) end and extends downwards by J octaves, with wavelet center frequencies

$$f_{j,q} = \frac{f_s}{2} 2^{-(j-1)-q/Q}, \quad (\text{B2})$$

which produces a logarithmic tiling of the frequency domain. Larger j thus corresponds to a coarser dyadic scale, a lower centre frequency, and a broader temporal support.

The width parameter $\sigma_{j,q}$ sets the temporal width of the Gaussian envelope, chosen as

$$\sigma_{j,q}(f_{j,q}, Q) = \frac{1}{f_{j,q}(2^{1/Q} - 2^{-1/Q})}, \quad (\text{B3})$$

which ensures an approximately constant relative bandwidth across scales (a constant- Q filter bank).

The WST is constructed using complex Morlet wavelets $\psi_{j,q}(t)$, localized in both time and frequency, given by

$$\psi_{j,q}(t) \approx \frac{1}{\sigma_{j,q}\sqrt{2\pi}} (e^{i2\pi f_{j,q}t}) e^{-t^2/(2\sigma_{j,q}^2)}, \quad (\text{B4})$$

The characteristic frequency associated with octave j is

$$\nu_j = \frac{f_s}{2^j}. \quad (\text{B5})$$

This construction yields a constant- Q filter bank spanning the physically accessible range

$$\frac{f_s}{2^J} \lesssim f_{j,q} \lesssim \frac{f_s}{2}, \quad (\text{B6})$$

where the upper bound is the Nyquist frequency and the lower bound is set by the maximum scale J , i.e., by how many octaves below Nyquist the bank extends. Increasing j moves to coarser scales and lower center frequencies, while Q fixes the number of wavelets per octave and hence the frequency resolution of the tiling.

In the time domain, each wavelet has a characteristic temporal support $\Delta t_{j,q}$ inversely proportional to its bandwidth,

$$\Delta t_{j,q} \sim \frac{1}{\Delta f_{j,q}} \sim \frac{Q}{f_{j,q}}, \quad (\text{B7})$$

so that high-frequency wavelets probe rapid oscillations while low-frequency wavelets capture slower modulations of the signal [73]. The parameter Q therefore controls the time–frequency trade-off of the transform: larger Q provides finer frequency resolution within each octave, while smaller Q improves temporal localization.

In practice, J is chosen such that

$$\tau_J \lesssim T, \quad (\text{B8})$$

ensuring that the averaging window does not exceed the duration of the analyzed strain segment. In this work, detector data segments have duration $T = 8$ s. At a sampling frequency of 2048 Hz, choosing $J = 7$ sets $\tau_J = 1/16$ s, with a corresponding lowest wavelet frequency $\nu_J = f_s/2^J = 16$ Hz.

APPENDIX C: 1D-CNN VS LR AND 2D-CNN

To set a baseline for the 1D-CNN, we apply a simpler LR classifier to the WST data. Our experiments show that an LR classifier alone is already sufficient to identify eccentricity in the data, achieving an AUC of ~ 0.8 and a TPR of ~ 0.55 at a target FPR of 10%. In addition we also compare against a more complex 2D-CNN classifier that operates directly on the two-dimensional WST representation. Its input consists of the logarithmically rescaled⁶ WST coefficient tensors, with convolutional kernels acting along both the temporal (T_b) and channel (C) dimensions, and with the three detectors treated as separate input channels, analogous to the RGB channels in image analysis. The network is designed to resemble the `ecc_sepcnn` model of Ref. [31] and is implemented in PyTorch [74]. Its core consists of separable convolutional layers followed by fully connected layers; a schematic overview is shown in Fig. 9. The number of channels in successive convolutional blocks increases as $32 \rightarrow 64 \rightarrow 128 \rightarrow 128 \rightarrow 256$, while the

⁶The raw WST coefficients span several orders of magnitude, which can adversely affect neural-network training.

TABLE IV. Performance of the alternative classifiers, the 2D-CNN and the LR. Performance is summarized by the threshold-independent metrics AUC (area under the ROC curve) and AP (average precision), evaluated on the test set for the best fold; the TPR is additionally reported at a target FPR of 10%. The different 2D-CNN setups are detailed in the text. Bold marks the best-performing alternative classifier in each column, the 2D-CNN with logarithmic rescaling.

Model	AUC	AP	TPR _{0.1}
Log Reg	0.815	0.847	0.562
2D-CNN			
Log-scale	0.841	0.872	0.630
Raw	0.814	0.849	0.563
Large	0.826	0.858	0.580

kernel sizes decrease as $5 \rightarrow 5 \rightarrow 5 \rightarrow 3 \rightarrow 3$. A dropout rate of 0.2 is used to mitigate overfitting, and the fully connected layers reduce the feature dimension as $64 \rightarrow 32 \rightarrow 16$, followed by a single output node. With these settings, the model contains 86369 trainable parameters.

In its default configuration, the best-performing fold achieves a test AUC of 0.841 after 24 epochs, with a TPR of 63.0% at a target FPR of 10%. As reported in Table IV, omitting the logarithmic rescaling of the WST coefficients (*Raw*) degrades performance, confirming the importance of dynamic-range compression for effective feature extraction. Doubling the number of trainable parameters (*Large*) yields no further gains, indicating that the baseline architecture already captures the relevant structure in the data.

Comparing all three classifiers in Tables II and IV, the 2D-CNN attains performance comparable to the 1D-CNN at a similar parameter count (Figs. 7 and 8 provide additional information on the performance of the 1D-CNN classifier), but at a substantially higher training cost. The LR and 1D-CNN are trained on a MacBook Pro (M3 chip), requiring ~ 12 min (including hyperparameter optimization) and ~ 10 min for 10-fold cross-validation, respectively. The 2D-CNN, by contrast, is trained on the CSD3 cluster (Ampere nodes, Cambridge, UK) using one A100 GPU and 32 CPU cores, incurring significantly greater computational overhead for equivalent predictive performance. This near-equivalence suggests that

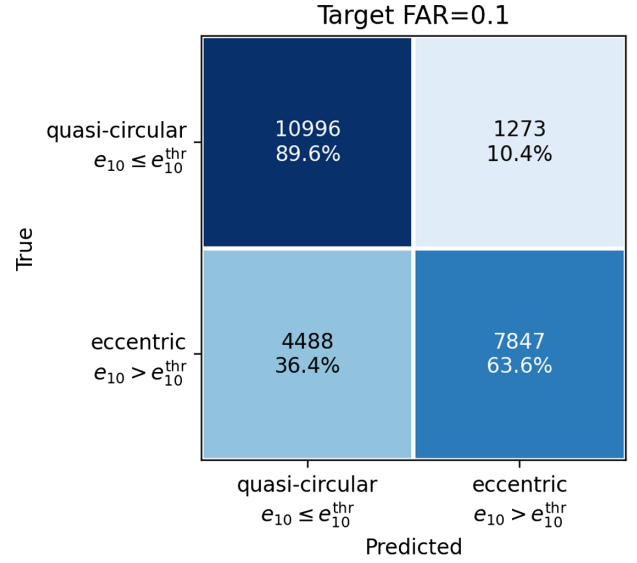


FIG. 7. Confusion matrix obtained for the reference 1D-CNN model and $J = 7$, $Q = 2$ WST coefficients, tuned to a target FPR of 10% on the validation set.

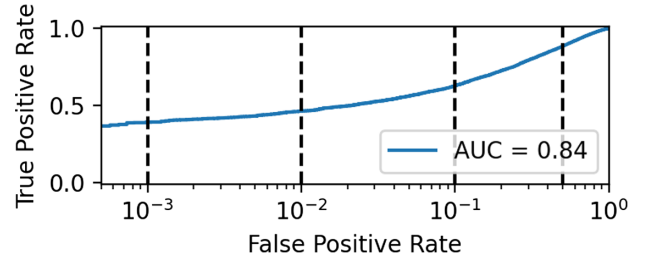


FIG. 8. ROC curve and AUC on the test set, for the best fold during training of the 1D-CNN model with $J = 7$, $Q = 2$ WST coefficients.

preserving the two-dimensional structure of the WST coefficients confers no significant advantage: unlike conventional time–frequency representations such as Q -transforms, the WST coefficients do not require an image-based interpretation, and their informative content can be captured effectively by one-dimensional models at substantially lower computational cost.

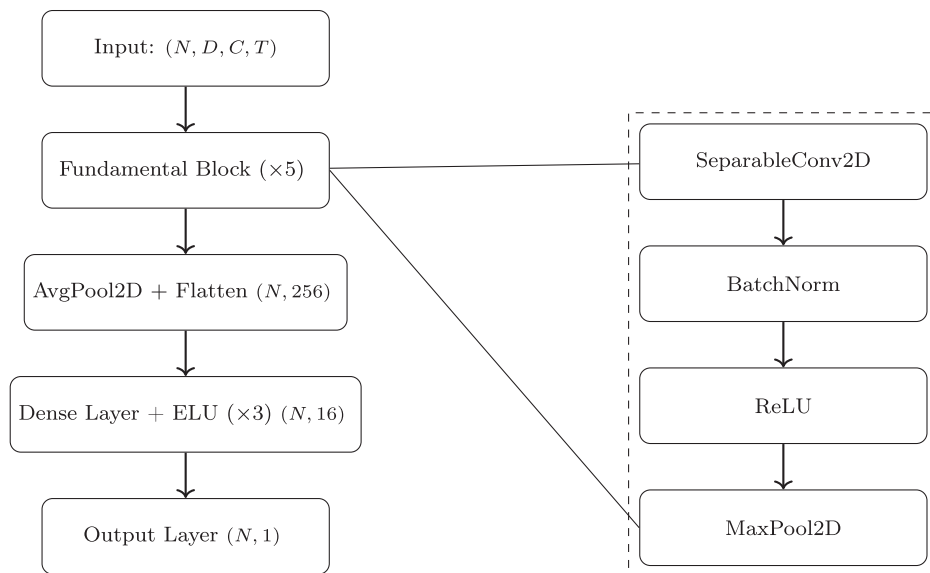


FIG. 9. Two-component diagram of SWT_CNN_2D. Left: the high-level architecture, showing the input, the fundamental layer (5× Fundamental Block), global pooling, the dense layers, and the output. Right: a close-up of a single Fundamental Block (SeparableConv2D + BatchNorm + ReLU + MaxPool). Tensor shapes are denoted (N, D, C, T) .

- [1] A. G. Abac *et al.*, GWTC-5.0: Observations from the second part of the fourth LIGO-Virgo-KAGRA observing run and updates to the gravitational-wave transient catalog, [arXiv:2605.27225](#).
- [2] A. G. Abac *et al.*, GWTC-5.0: Population properties of merging compact binaries, [arXiv:2605.27226](#).
- [3] J. Stegmann, F. Antonini, A. Olejak, S. Biscoveanu, V. Raymond, S. Rinaldi, and E. Flanagan, Gravitational-wave observations suggest most black hole mergers form in triples, *Astrophys. J. Lett.* **1000**, L59 (2026).
- [4] F. Antonini, I. Romero-Shaw, T. Callister, F. Dosopoulou, D. Chattopadhyay, Y. B. Ginat, M. Gieles, and M. Mapelli, Gravitational waves reveal the pair-instability mass gap and constrain nuclear burning in massive stars, *Nat. Astron.* **10**, 1049 (2026).
- [5] H. Tong, T. A. Callister, M. Fishbach, E. Thrane, F. Antonini, S. Stevenson, I. M. Romero-Shaw, and F. Dosopoulou, A subpopulation of low-mass, spinning black holes: signatures of dynamical assembly, [arXiv:2511.05316](#).
- [6] P. C. Peters, Gravitational radiation and the motion of two point masses, *Phys. Rev.* **136**, B1224 (1964).
- [7] F. Antonini, S. Toonen, and A. S. Hamers, Binary black hole mergers from field triples: Properties, rates, and the impact of stellar evolution, *Astrophys. J.* **841**, 77 (2017).
- [8] J. Samsing, Eccentric black hole mergers forming in globular clusters, *Phys. Rev. D* **97**, 103014 (2018).
- [9] I. Romero-Shaw, P. D. Lasky, and E. Thrane, Four eccentric mergers increase the evidence that LIGO-Virgo-KAGRA's binary black holes form dynamically, *Astrophys. J.* **940**, 171 (2022).
- [10] N. Gupte, A. Ramos-Buades, A. Buonanno, J. Gair, M. C. Miller, M. Dax, S. R. Green, M. Pürrer, J. Wildberger, J. Macke, I. M. Romero-Shaw, and B. Schölkopf, Evidence for eccentricity in the population of binary black holes observed by LIGO-Virgo-KAGRA, *Phys. Rev. D* **112**, 104045 (2025).
- [11] M. d. L. Planas, A. Ramos-Buades, C. García-Quirós, H. Estellés, S. Husa, and M. Haney, Reanalysis of binary black hole gravitational wave events for orbital eccentricity signatures, *Phys. Rev. D* **112**, 123004 (2025).
- [12] G. Morras, G. Pratten, and P. Schmidt, Orbital eccentricity in a neutron star-black hole merger, *Astrophys. J. Lett.* **1000**, L2 (2026).
- [13] Y. Xu, J. Valencia, H. Estellés Estrella, A. Ramos Buades, S. Husa, M. Rosselló-Sastre, J. Llobera Querol, F. Ramis Vidal, M. de Lluc Planas Llompart, M. Colleoni, E. Hamilton, A. Montava Agudo, J. Yébana Carrilero, and A. Heffernan, Parameter estimation for the GWTC-4.0 catalog with phenomenological waveform models that include orbital eccentricity and an updated description of spin precession, *Phys. Rev. D* **113**, 083001 (2026).
- [14] K. Paul, A. Maurya, Q. Henry, K. Sharma, P. Satheesh, Divyajyoti, P. Kumar, and C. K. Mishra, Eccentric, spinning, inspiral-merger-ringdown waveform model with higher modes for the detection and characterization of binary black holes, *Phys. Rev. D* **111**, 084074 (2025).
- [15] K. S. Phukon, P. Schmidt, and G. Pratten, Geometric template bank for the detection of spinning low-mass compact binaries with moderate orbital eccentricity, *Phys. Rev. D* **111**, 043040 (2025).

- [16] I. M. Romero-Shaw, D. Gerosa, and N. Loutrel, Eccentricity or spin precession? Distinguishing subdominant effects in gravitational-wave data, *Mon. Not. R. Astron. Soc.* **519**, 5352 (2023).
- [17] K. Divyajyoti, Sumit, S. Tibrewal, I. M. Romero-Shaw, and C. K. Mishra, Blind spots and biases: The dangers of ignoring eccentricity in gravitational-wave signals from binary black holes, *Phys. Rev. D* **109**, 043037 (2024).
- [18] Divyajyoti, I. M. Romero-Shaw, V. Prasad, K. Paul, C. Kant Mishra, P. Kumar, A. Maurya, M. Boyle, L. E. Kidder, H. P. Pfeiffer, and M. A. Scheel, Biased parameter inference of eccentric, spin-precessing binary black holes, *Phys. Rev. D* **113**, 103022 (2026).
- [19] S. Tibrewal, A. Zimmerman, J. Lange, and D. Shoemaker, Misinterpreting spin precession as orbital eccentricity in gravitational-wave signals, [arXiv:2601.02260](https://arxiv.org/abs/2601.02260).
- [20] E. Payne, S. Hourihane, J. Golomb, R. Udall, D. Davis, and K. Chatziioannou, Curious case of GW200129: Interplay between spin-precession inference and data-quality issues, *Phys. Rev. D* **106**, 104017 (2022).
- [21] I. Romero-Shaw, J. Stegmann, H. Tagawa, D. Gerosa, J. Samsing, N. Gupte, and S. R. Green, GW200208_222617 as an eccentric black-hole binary merger: Properties and astrophysical implications, *Phys. Rev. D* **112**, 063052 (2025).
- [22] M. Zevin, I. M. Romero-Shaw, K. Kremer, E. Thrane, and P. D. Lasky, Implications of eccentric observations on binary black hole formation channels, *Astrophys. J. Lett.* **921**, L43 (2021).
- [23] D. A. Brown and P. J. Zimmerman, Effect of eccentricity on searches for gravitational waves from coalescing compact binaries in ground-based detectors, *Phys. Rev. D* **81**, 024007 (2010).
- [24] B. Gadre, K. Soni, S. Tiwari, A. Ramos-Buades, M. Haney, and S. Mitra, Detectability of eccentric binary black holes with matched filtering and unmodeled pipelines during the third observing run of LIGO-Virgo-KAGRA, *Phys. Rev. D* **110**, 044013 (2024).
- [25] A. Ramos-Buades, S. Tiwari, M. Haney, and S. Husa, Impact of eccentricity on the gravitational-wave searches for binary black holes: High mass case, *Phys. Rev. D* **102**, 043005 (2020).
- [26] I. Romero-Shaw, J. Stegmann, G. Morras, A. Dorozsmai, and M. Zevin, Astrophysical implications of eccentricity in gravitational waves from neutron star-black hole binaries, *Mon. Not. R. Astron. Soc.* **547**, stag323 (2026).
- [27] I. M. Romero-Shaw, P. D. Lasky, and E. Thrane, Searching for eccentricity: Signatures of dynamical formation in the first gravitational-wave transient catalogue of LIGO and Virgo, *Mon. Not. R. Astron. Soc.* **490**, 5210 (2019).
- [28] G. Morras, G. Pratten, and P. Schmidt, Improved post-Newtonian waveform model for inspiralling precessing-eccentric compact binaries, *Phys. Rev. D* **111**, 084052 (2025).
- [29] A. Gamboa, A. Buonanno, R. Enficiaud, M. Khalil, A. Ramos-Buades, L. Pompili, H. Estellés, M. Boyle, L. E. Kidder, H. P. Pfeiffer, H. R. Rüter, and M. A. Scheel, Accurate waveforms for eccentric, aligned-spin binary black holes: The multipolar effective-one-body model SEOBNRv5EHM, *Phys. Rev. D* **112**, 044038 (2025).
- [30] S. Albanesi, R. Gamba, S. Bernuzzi, J. Fontbuté, A. Gonzalez, and A. Nagar, Effective-one-body modeling for generic compact binaries with arbitrary orbits, *Phys. Rev. D* **112**, L121503 (2025).
- [31] A. Ravichandran, A. Vijaykumar, S. J. Kapadia, and P. Kumar, Rapid identification and classification of eccentric gravitational wave inspirals with machine learning, [arXiv:2302.00666](https://arxiv.org/abs/2302.00666).
- [32] E. Khalouei, C. G. Sabiu, H. M. Lee, and A. Gopakumar, External attention transformer: A robust AI model for identifying initial eccentricity signatures in binary black hole events in simulated Advanced LIGO data, *J. Cosmol. Astropart. Phys.* **10** (2025) 028.
- [33] S. Mallat, Group invariant scattering, *Pure Appl. Math.* **65**, 1331 (2012).
- [34] J. Bruna and S. Mallat, Invariant scattering convolution networks, *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 1872 (2013).
- [35] A. Licciardi, D. Carbone, L. Rondoni, and A. Nagar, Wavelet scattering transform for gravitational wave analysis: An application to glitch characterization, *Phys. Rev. D* **111**, 084044 (2025).
- [36] S. Cheng and B. Ménard, How to quantify fields or textures? A guide to the scattering transform, [arXiv:2112.01288](https://arxiv.org/abs/2112.01288).
- [37] S. Cheng, Y.-S. Ting, B. Ménard, and J. Bruna, A new approach to observational cosmology using the scattering transform, *Mon. Not. R. Astron. Soc.* **499**, 5902 (2020).
- [38] G. Valogiannis and C. Dvorkin, Towards an optimal estimation of cosmological parameters with the wavelet scattering transform, *Phys. Rev. D* **105**, 103534 (2022).
- [39] G. Valogiannis, F. Villaescusa-Navarro, and M. Baldi, Towards unveiling the large-scale nature of gravity with the wavelet scattering transform, *J. Cosmol. Astropart. Phys.* **11** (2024) 061.
- [40] H. Shimabukuro, Y. Xu, and Y. Shao, Analyzing the 21-cm forest with wavelet scattering transform: Insight into non-Gaussian features of the 21-cm forest, *Phys. Rev. D* **112**, 063557 (2025).
- [41] R. Morel, G. Rochette, R. Leonarduzzi, J.-P. Bouchaud, and S. Mallat, Scale dependencies and self-similar models with wavelet scattering spectra, *Appl. Comput. Harmon. Anal.* **75**, 101724 (2025).
- [42] R. Morel, S. Mallat, and J.-P. Bouchaud, Path shadowing Monte Carlo, *Quant. Finance* **24**, 1199 (2024).
- [43] S. Cheng, R. Morel, E. Allys, B. Ménard, and S. Mallat, Scattering spectra models for physics, *PNAS Nexus* **3**, pgae103 (2024).
- [44] T. Angles and S. Mallat, Generative networks as inverse problems with scattering transforms (ICLR, 2018).
- [45] A. Siahkoohi, R. Morel, M. V. de Hoop, E. Allys, G. Sainton, and T. Kawamura, Unearthing insights into mars: Unsupervised source separation with limited data, in *Proceedings of the 40th International Conference on Machine Learning* (PMLR, 2023).
- [46] G. Ashton, M. Hübner, P. D. Lasky, C. Talbot, K. Ackley, S. Biscoveanu, Q. Chu, A. Divakarla, P. J. Easter, B. Goncharov, F. Hernandez Vivanco, J. Harms, M. E. Lower, G. D. Meadors, D. Melchor, E. Payne, M. D. Pitkin, J. Powell, N. Sarin, R. J. E. Smith, and E. Thrane, Bilby: A user-friendly Bayesian inference library for

- gravitational-wave astronomy, *Astrophys. J. Suppl. Ser.* **241**, 27 (2019).
- [47] I. M. Romero-Shaw *et al.*, Bayesian inference for compact binary coalescences with Bilby: Validation and application to the first LIGO-Virgo gravitational-wave transient catalogue, *Mon. Not. R. Astron. Soc.* **499**, 3295 (2020).
- [48] A. Vijaykumar, A. G. Hanselman, and M. Zevin, Consistent eccentricities for gravitational-wave astronomy: Resolving discrepancies between astrophysical simulations and waveform models, *Astrophys. J.* **969**, 132 (2024).
- [49] M. A. Shaikh, V. Varma, H. P. Pfeiffer, A. Ramos-Buades, and M. van de Meent, Defining eccentricity for gravitational wave astronomy, *Phys. Rev. D* **108**, 104007 (2023).
- [50] J. Anden and S. Mallat, Deep scattering spectrum, *IEEE Trans. Signal Process.* **62**, 4114 (2014).
- [51] M. Andreux, T. Angles, G. Exarchakis, R. Leonarduzzi, G. Rochette, L. Thiry, J. Zarka, S. Mallat, J. Andén, E. Belilovsky, J. Bruna, V. Lostanlen, M. Chaudhary, M. J. Hirn, E. Oyallon, S. Zhang, C. Cella, and M. Eickenberg, Kymatio: Scattering transforms in Python, *Mach. Learn. Res.* **21**, 1 (2020), <http://jmlr.org/papers/v21/19-047.html>.
- [52] B. G. Patterson, S. M. Tomson, and S. Fairhurst, Identifying eccentricity in binary black hole mergers using a harmonic decomposition of the gravitational waveform, *Phys. Rev. D* **111**, 044073 (2025).
- [53] T. Islam, G. Khanna, and S. E. Field, Adding higher-order spherical harmonics in non-spinning eccentric binary black hole merger waveform models, *Phys. Rev. D* **111**, 124023 (2025).
- [54] J. Bruna and S. Mallat, Classification with invariant scattering representations, in *2011 IEEE 10th IVMSP Workshop: Perception and Visual Signal Analysis* (IEEE, Ithaca, NY, 2011), pp. 99–104.
- [55] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, SCIKIT-LEARN: Machine learning in Python, *J. Mach. Learn. Res.* **12**, 2825 (2011), <http://jmlr.org/papers/v12/pedregosa11a.html>.
- [56] H. Tang, J. Yang, B. Wang, and T. Yang, Impact of higher-order modes on eccentricity measurement in binary black hole gravitational waves, *Phys. Rev. D* **114**, 044027 (2026).
- [57] A. Ramos-Buades, A. Buonanno, H. Estellés, M. Khalil, D. P. Mihaylov, S. Ossokine, L. Pompili, and M. Shiferaw, SEOBNRv5PHM: Next generation of accurate and efficient multipolar precessing-spin effective-one-body waveforms for binary black holes, *Phys. Rev. D* **108**, 124037 (2023).
- [58] LVK Collaboration (private communication).
- [59] P. Canizares, S. J. Staelens, and I. Romero-Shaw, WST-Eccentricity, <https://github.com/Priscilla/wst-eccentricity> (2026), Accessed: 2026-07-07.
- [60] <https://zenodo.org/records/21204498>.
- [61] K. S. Thorne, Multipole expansions of gravitational radiation, *Rev. Mod. Phys.* **52**, 299 (1980).
- [62] L. E. Kidder, Using full information when computing modes of post-Newtonian waveforms from inspiralling compact binaries in circular orbit, *Phys. Rev. D* **77**, 044016 (2008).
- [63] A. Buonanno and T. Damour, Effective one-body approach to general relativistic two-body dynamics, *Phys. Rev. D* **59**, 084006 (1999).
- [64] T. Damour, P. Jaranowski, and G. Schäfer, On the determination of the last stable orbit for circular general relativistic binaries at the third post-Newtonian approximation, *Phys. Rev. D* **62**, 084011 (2000).
- [65] Y. Pan, A. Buonanno, M. Boyle, L. T. Buchman, L. E. Kidder, H. P. Pfeiffer, and M. A. Scheel, Inspiral-merger-ringdown multipolar waveforms of nonspinning black-hole binaries using the effective-one-body formalism, *Phys. Rev. D* **84**, 124052 (2011).
- [66] S. Ossokine *et al.*, Multipolar effective-one-body waveforms for precessing binary black holes: Construction and validation, *Phys. Rev. D* **102**, 044055 (2020).
- [67] T. Damour, B. R. Iyer, and A. Nagar, Improved resummation of post-Newtonian multipolar waveforms from circularized compact binaries, *Phys. Rev. D* **79**, 064004 (2009).
- [68] T. Hinderer and S. Babak, Foundations of an effective-one-body model for coalescing binaries on eccentric orbits, *Phys. Rev. D* **96**, 104048 (2017).
- [69] D. Chiamarello and A. Nagar, Faithful analytical effective-one-body waveform model for spin-aligned, moderately eccentric, coalescing black hole binaries, *Phys. Rev. D* **101**, 101501 (2020).
- [70] S. Babak, A. Taracchini, and A. Buonanno, Validating the effective-one-body model of spinning, precessing binary black holes against numerical relativity, *Phys. Rev. D* **95** (2017).
- [71] P. Schmidt, F. Ohme, and M. Hannam, Towards models of gravitational waveforms from generic binaries. II. Modelling precession effects with a single effective precession parameter, *Phys. Rev. D* **91**, 024043 (2015).
- [72] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, Simple model of complete precessing black-hole-binary gravitational waveforms, *Phys. Rev. Lett.* **113**, 151101 (2014).
- [73] J. Andén, V. Lostanlen, and S. Mallat, Joint time-frequency scattering for audio classification, in *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)* (IEEE, New York, 2015), pp. 1–6.
- [74] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimeshine, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, *PyTorch: An imperative style, high-performance deep learning library* (Curran Associates, Inc., Red Hook, NY, 2019).