DANSur: Dual-Stage Artificial Neural Surrogate with Higher Modes

Speaker: Osvaldo Gramaxo Freitas

Universitat de València, Spain · CF-UM-UP, Portugal

Co-authors: A. Theodoropoulos, N. Villanueva, J. A. Font, A. Onofre, A. Torres-Forné, J. D. Martín-Guerrero

CoCoNuT Meeting 2026
University of Southampton
8 September 2026

This work is supported by the Spanish Agencia Estatal de Investigación (grant PID2024-159689NB-C21) funded by the Ministerio de Ciencia, Innovación y Universidades.

Gravitational Wave Astronomy Requirements

Observational Era

- **100+ Detections:** Compact binary coalescences observed across LIGO, Virgo, and KAGRA.
- **Next-Generation Observatories:** Einstein Telescope (ET), Cosmic Explorer (CE), and LISA.
- $\mathbf{10^4 - 10^5}$ **Events/Year** expected.

The Computational Challenge

- Bayesian parameter estimation (PE) evaluates likelihood:

$$\ln \mathcal{L}(\mathbf{d} \mid \boldsymbol{\theta}) \propto -\frac{1}{2}\langle \mathbf{d} - h(\boldsymbol{\theta}) \mid \mathbf{d} - h(\boldsymbol{\theta}) \rangle$$

- Requires $\mathbf{10^6 - 10^7}$ **waveform evaluations** per event.
- Standard pipelines take **days to weeks per event**.

GW approximants

Numerical Relativity

- Solves Einstein's field equations $G_{\mu\nu} = 8\pi T_{\mu\nu}$ exactly.
- **Gold standard** fidelity.
- **Prohibitive cost:** $10^4 - 10^5$ CPU-hours/run.
- Only $\sim \mathcal{O}(10^3)$ total simulations exist (SXS).

Approximants (EOB)

- Effective-One-Body (`SEOBNR`) and Phenom models.
- Tuned semi-analytic ODEs.
- Generation time $\sim 0.1 - 1$ s.
- Purely CPU-bound; too slow for modern MCMC / nested sampling.

Standard Surrogates

- E.g., `NRHybSur3dq8`.
- Interpolation over reduced bases.
- High accuracy, but:
  - Memory-intensive splines.
  - No GPU acceleration.
  - **Non-differentiable**.

The Solution: An ultra-fast, GPU-accelerated, **natively differentiable** surrogate trained directly on Numerical Relativity.

Higher Multipole Modes

Waveform polarizations expand into spin-weighted spherical harmonics ${}_{-2}Y_{lm}(\theta, \phi)$:

$$h_+(t) - ih_\times(t) = \sum_{l=2}^{\infty} \sum_{m=-l}^{l} h_{lm}(t)\, {}_{-2}Y_{lm}(\theta, \phi)$$

Why Higher Modes Matter

- Dominant $(2, \pm 2)$ mode carries most radiated power.
- Higher modes $(3,3), (2,1), (4,4), \ldots$ are amplified in **asymmetric binaries** ($q = m_1/m_2 > 1$) and **inclined systems**.
- Crucial for breaking the distance-inclination ($d_L - \iota$) degeneracy

The DANSur Philosophy: Dual-Stage Transfer Learning

STAGE 1: Dense Pre-Training

$\sim 10^5$ synthetic waveforms from base approximant (`NRHybSur3dq8`)

$\Longrightarrow$ Establishes smooth global parameter manifold.

Transfer $\longrightarrow$

STAGE 2: NR Fine-Tuning

$N = 264$ Numerical Relativity simulations (SXS Catalog)

$\Longrightarrow$ Adapts weights to true* physics.

- Neural networks require dense coverage to prevent unphysical interpolation artifacts.
- Approximants provide infinite training waveforms across $(q, \chi_{1z}, \chi_{2z})$.
- Note: previous work (doi.org/10.1103/7bkx-hs53) has shown that differently accurate base approximants still lead to similar results after fine-tuning

Dimensionality Reduction: Amplitude-Phase PCA

Polar Waveform Representation

Decompose into physical observables:

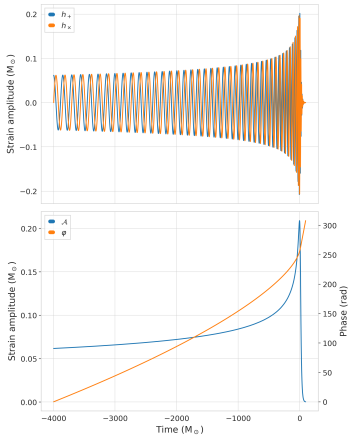
$$h_{lm}(t) = A_{lm}(t) \exp\left(-i\,\phi_{lm}(t)\right)$$

- $A_{lm}(t) \geq 0$: Smooth bell-shaped envelope.
- $\phi_{lm}(t)$: Monotonically increasing orbital phase.
- Avoids oscillatory cancellation of complex $h(t)$.

Linear PCA Compression

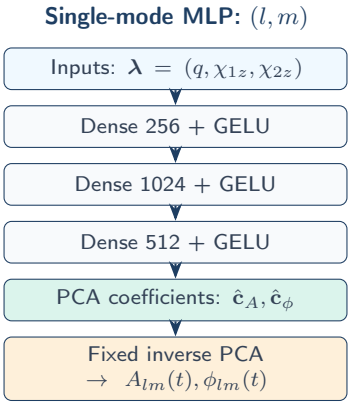
$$A_{lm}(t) = \mu_A(t) + \sum_{k=1}^{40} \hat{c}_{A,k} V_k^{(A)}(t),$$

$$\phi_{lm}(t) = \mu_\phi(t) + \sum_{k=1}^{45} \hat{c}_{\phi,k} V_k^{(\phi)}(t).$$



Clean decomposition into envelope $A(t)$ and unwrapped phase $\phi(t)$.

From Binary Parameters to PCA Coefficients



Predict a Compact Waveform Representation

Each mode learns $f_{lm}(\boldsymbol{\lambda}) = (\hat{\mathbf{c}}_A, \hat{\mathbf{c}}_\phi)$.

- Dense+GELU layers learn nonlinear dependence on mass ratio and spins.
- Output: **40 amplitude + 45 phase coefficients**, rather than full time series.
- Coefficient targets come from PCA of training waveforms; waveform losses also guide training.
- Small overall network size allows for very fast inference!

The Challenge with Higher Modes: Phase Singularities

Equatorial and Orbital Symmetry

For aligned-spin binaries, reflection symmetry enforces:

$$h_{lm}(t) \equiv 0 \quad \text{for odd } m \quad \text{when } q = 1 \text{ and } \chi_{1z} = \chi_{2z}$$

The Phase Singularity Problem

When amplitude vanishes ($A_{lm} \to 0$):

$$\phi_{lm}(t) = \mathrm{Arg}(0) \quad \text{is mathematically undefined}$$

- Mismatch metric becomes undefined:

$$\mathfrak{M}(h_1, h_2) = 1 - \frac{\langle h_1,\, h_2 \rangle}{\sqrt{\langle h_1, h_1 \rangle \, \langle h_2, h_2 \rangle}}\,, \tag{1}$$

- Destabilizes neural network training if mismatch is used by itself.

Power-Weighted Loss

Power-Weighted Mismatch Loss (Odd-m)

For odd-m modes, each waveform i receives one scalar weight from its integrated power:

$$P_i = \int \left| h^{\text{true}}_{lm,i}(t) \right|^2 \, dt, \qquad \omega_i = \frac{P_i}{\max_{j \in \mathcal{B}} P_j},$$

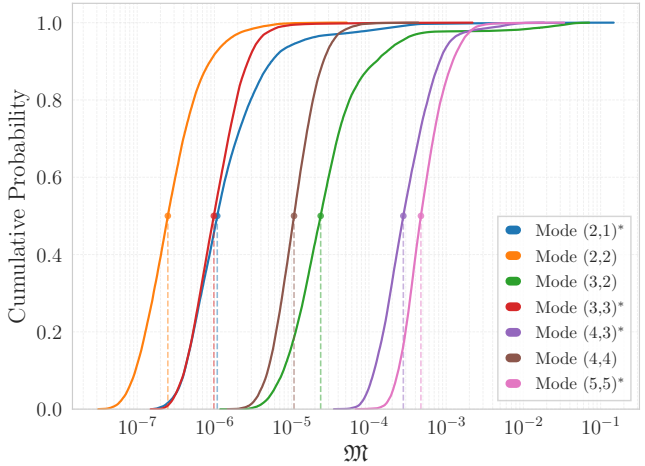
where $\mathcal{B}$ is the training batch

$$\mathcal{L}_{\text{pretrain}} = \log_{10}\left(\langle \omega_i \, \mathfrak{M}_i \rangle_{i \in \mathcal{B}}\right) + 10\,\text{MAE}(\mathbf{c}, \hat{\mathbf{c}}) + \langle |P_{\text{true}} - P_{\text{pred}}| \rangle,$$

Physical and Numerical Behavior

- **Low power relative to the batch maximum:** Small ω_i suppresses the waveform's mismatch contribution when its odd-m mode is weak.
- **Maximum-power waveform:** $\omega_i = 1$. Its mismatch retains full weight; other waveforms are weighted by their relative integrated power.

Power-Weighted Loss



Spherical Shell Loss and PN Matching

1. Global Spherical Shell Loss

Individual mode losses do not guarantee coherent total strain. We project onto the sphere S^2 to cover inclination and orbital phase for SXS waveforms:

$$\mathcal{L}_{S^2} = \log_{10}\left(\frac{1}{K}\sum_{k=1}^{K}\mathcal{M}\Big(h^{\text{pred}}(\theta_k,\phi_k),\, h^{\text{true}}\Big)\right)$$

- Evaluated over 20×20 sphere points.
- Enforces physical multi-mode interference.

2. Post-Newtonian Gauge Fixing

Fixes relative inter-mode phase offsets ξ_{lm} at initial frequency $v_0 = (\omega_{22}/2)^{1/3}$:

$$\xi_{lm} = \text{Arg}\left(H^{\text{PN}}_{lm}(v_0; q, \chi_{1z}, \chi_{2z})\right)$$

- Evaluated analytically in $< 1\,\mu$s via Numba.
- Guarantees seamless phase consistency with PN inspiral without free parameters.

Inference Architecture

Overcoming Multi-Model Overhead

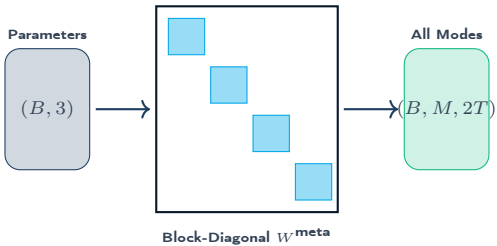
Evaluating 4–7 separate mode decoders sequentially or via multiple streams creates severe GPU kernel launch overhead.

The MetaNet Architecture

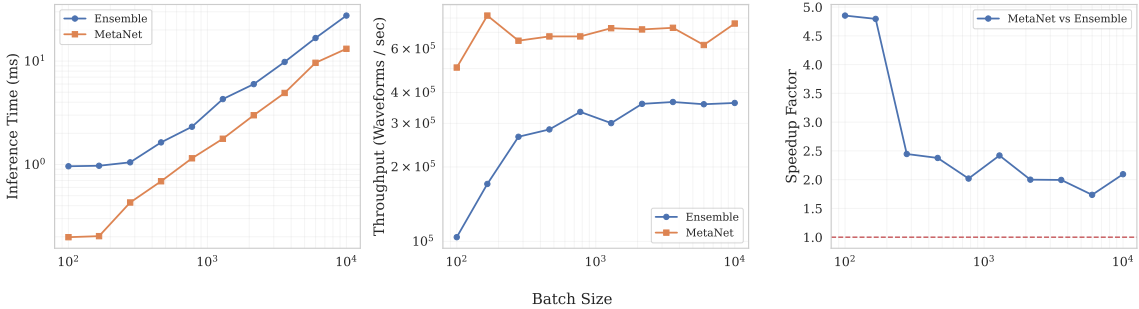
- **Layer 1:** Stacks weights vertically into a single GEMM.
- **Hidden Layers:** Block-diagonal matrix structure:

$$W_l^{\text{meta}} = \text{diag}\left(W_l^{(2,2)}, W_l^{(3,3)}, \dots\right)$$

- **Inverse PCA:** Single batched tensor contraction: `torch.einsum('bij,ijk->bik')`



Inference Speed and Hardware Benchmarks



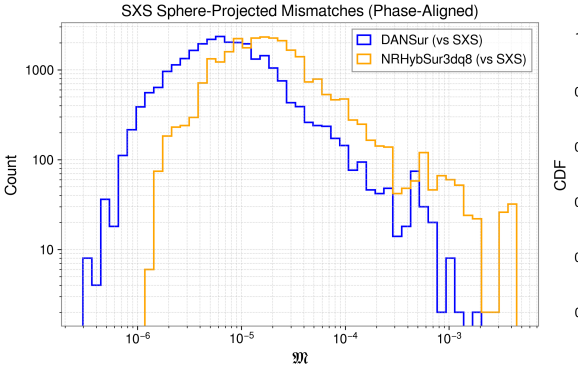
Latency, throughput, and speedup on NVIDIA V100 GPU.

Accuracy Validation: Benchmarking vs $N = 264$ SXS Simulations

Mismatch Metric	DANSur	NRHybSur3dq8
Median Sphere-Avg	$\mathbf{7.52 \times 10^{-6}}$	1.72×10^{-5}
Median Point-by-Point	$\mathbf{6.74 \times 10^{-6}}$	1.66×10^{-5}
Max Sphere-Avg	$\mathbf{4.95 \times 10^{-4}}$	1.96×10^{-3}
Max Point-by-Point	$\mathbf{1.94 \times 10^{-3}}$	4.50×10^{-3}

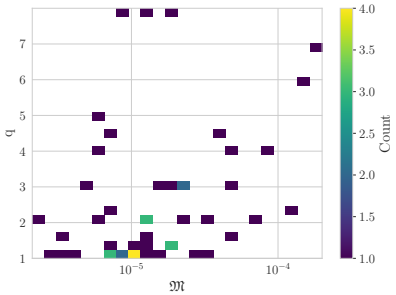
Key Highlights

- Evaluated against all $N = 264$ SXS NR runs.
- DANSur achieves $> 2\times$ **lower median mismatch** than `NRHybSur3dq8`.
- Worst-case error well below detector calibration thresholds ($\sim 10^{-3}$).

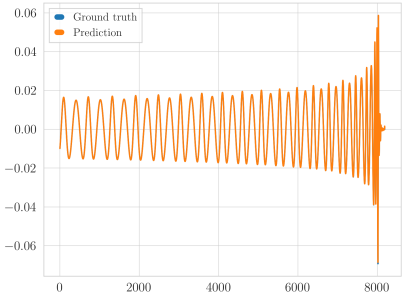


Distribution of sphere-projected mismatches vs SXS NR.

Parameter Space Coverage and Worst-Case Robustness



Average mismatch vs mass ratio $q \in [1, 8]$ across SXS.

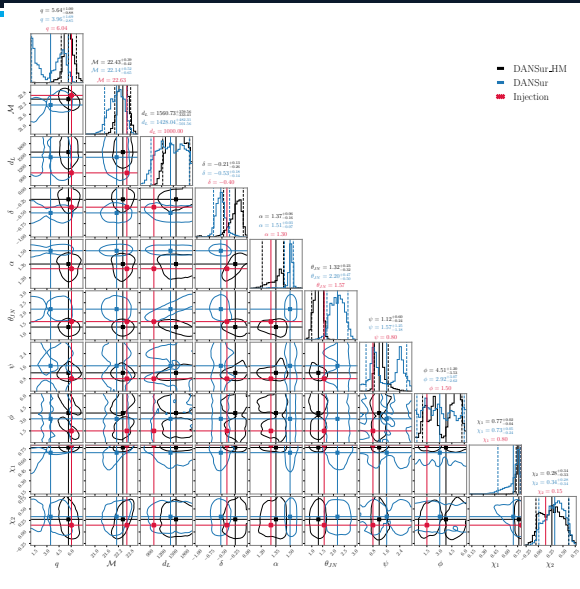


99.9th percentile worst-case reconstruction vs SXS NR.

Higher Modes: Effect on PE

SXS:BBH$_1$437

- Sanity test: PE with higher modes is expected to be more accurate than (2,2) mode only
- Injecting an NR event from the SXS catalog into noise ensures we know ground-truth of the data
- DANSur_HM indeed shows better parameter recovery than what is obtained with previous (2,2)-only model.



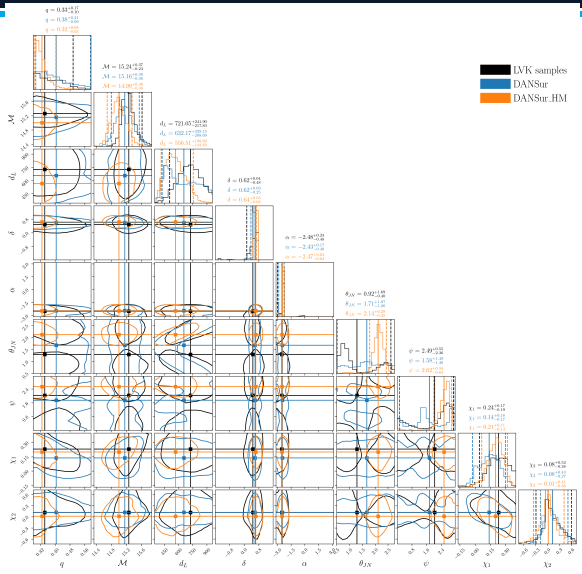
Application to Real Gravitational Wave Events

GW190412: Unequal-Mass Benchmark

- Asymmetric merger: $q \approx 3.6$ with primary spin $\chi_1 \approx 0.43$.
- First event with decisive observational detection of the $(3, 3)$ mode.
- DANSur reproduces LVK publication posteriors with high fidelity.

Other Validated Events

- **GW190814:** Extreme mass ratio $q \approx 9.2$.
- **GW230814:** High-SNR O4 detection with prominent higher harmonics.



Native Differentiability

The Finite-Difference Bottleneck

$$\frac{\partial h}{\partial \theta_i} \approx \frac{h(\boldsymbol{\theta} + \epsilon \mathbf{e}_i) - h(\boldsymbol{\theta} - \epsilon \mathbf{e}_i)}{2\epsilon}$$

- Requires $2N$ full waveform evaluations.
- Highly sensitive to step size ϵ (numerical cancellation vs truncation error).
- Non-differentiable on spline knots!

DANSur Forward-Mode AD (`torch.func.jvp`)

- Built entirely with native, smooth PyTorch/JAX operations (C^∞ GELU activations).
- Computes exact Jacobian-Vector Products:

$$J_v = \nabla_{\boldsymbol{\theta}} h(\boldsymbol{\theta}) \cdot \mathbf{v}$$

- **Single forward pass** with machine precision.
- Zero numerical noise, fully vectorized on GPU.

Native Differentiability

The Degeneracy Problem

For $(2, 2)$-only models, luminosity distance d_L and inclination ι share nearly identical detector projections:

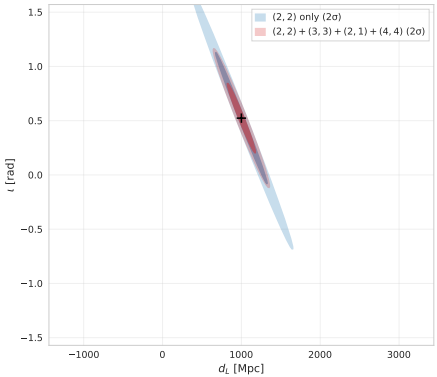
$$h_+ \propto \frac{1 + \cos^2 \iota}{2\, d_L}, \quad h_\times \propto \frac{\cos \iota}{d_L}$$

producing broad, degenerate posteriors.

The Higher-Mode Solution

Modes $(3, 3)$ and $(2, 1)$ scale with different inclination powers:

$${}_{-2}Y_{33} \propto \sin \iota (1 + \cos \iota)^2, \quad {}_{-2}Y_{21} \propto \sin \iota (1 + \cos \iota)$$



Fisher ellipses in (d_L, ι) comparing $(2, 2)$-only vs all higher modes.

Conclusions

- **Fidelity:** Multi-modal surrogate achieves average mismatches $\sim 10^{-4}$ across SXS NR data.
- **Throughput:** Meta-network generates $> 6 \times 10^5$ waveforms/s on an NVIDIA V100 GPU.
- **Power Weighting:** Solves the singular vanishing-amplitude phase problem for odd-m modes near $q = 1$.
- **Astrophysical Utility:** Fully eliminates parameter estimation biases in edge-on, asymmetric coalescences.

Future Directions and Community Release

Possible Future Extensions

- **Precession Modeling (7D/15D):** Incorporating in-plane spins into coprecessing frames.
- **Orbital Eccentricity:** Dynamical capture binaries.
- **Higher Mass Ratios ($q > 10$):** Intermediate Mass Ratio Inspirals (IMRIs) for LISA.
- **Production Integration:** Native plugins for `Bilby`, `PyCBC`, and `ripple`.

Open-Source Code Release

- Code available **github.com/osvaldogramaxo/DANSur_HM**, along with a pre-trained model and `gwsurrogate` compatible implementation
- Preprint available at arXiv:2609.03025(or scan QR!)

Thank You for Your Attention!

Backup: MetaNetwork Tensor Operations

Mathematical Structure

For $M = 4$ modes and batch size B:

1. **Linear Projection:**

$$\mathbf{z} = W_L^{\text{meta}} \dots \sigma\left(W_1^{\text{meta}} \boldsymbol{\theta} + \mathbf{b}_1^{\text{meta}}\right)$$

where $\mathbf{z} \in \mathbb{R}^{B \times (M \cdot n_{\text{latent}})}$.

2. **Reshape into Mode Latents:**

$$\mathbf{Z} = \text{reshape}(\mathbf{z}, [B, M, n_{\text{latent}}])$$

3. **Batched Inverse PCA (Einsum):**

$$A_{b,m,t} = \sum_j Z_{b,m,j} \, V^{(A)}_{m,j,t} + \bar{A}_{m,t}$$

GPU Memory and Kernel Efficiency

- Standard ensembles launch M separate GPU kernels.
- MetaNet executes a single GEMM and a single `einsum` kernel.
- Maximizes memory bandwidth saturation.
- Fully compatible with `torch.compile`.

Backup: Training Schedules and Hyperparameters

Network Architecture

- **Inputs:** Dimensionless $(q, \chi_{1z}, \chi_{2z})$.
- **Hidden layers:** 3 layers ($[256, 1024, 512]$ per mode).
- **Activation:** GELU.
- **Precision:** Single precision `float32`.

Optimization and Validation

- **Pre-training:** AdamW ($\text{lr}_0 = 10^{-3}$, cosine decay).
- **Fine-tuning:** AdamW (initial lr of 10^{-5})