Testing ultralight dark matter with pulsar timing arrays

Chris Gordon
with Shreyas Tiruvaskar

School of Physical and Chemical Sciences
University of Canterbury

CosPA 2026 / ACGRG 13

Main article: S. Tiruvaskar and C. Gordon, arXiv:2605.05679
Background article: S. Tiruvaskar, R. Boey, R. Easther and C. Gordon, Phys. Rev. D **113**, 063541 (2026)

Pulsar Timing Arrays in One Slide

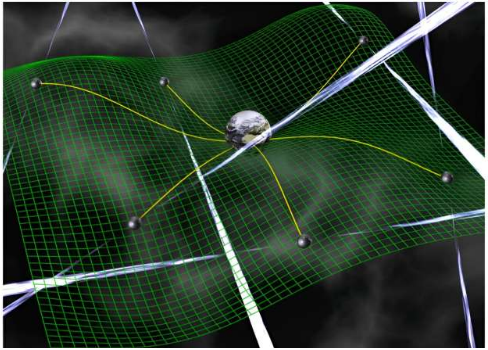


Image credit: David J. Champion / MPIfR

- Millisecond pulsars are extremely stable clocks.
- Nanohertz gravitational waves perturb pulse arrival times over years.
- A stochastic background produces correlated timing residuals across the sky.

Why this matters

PTA frequencies probe the slow inspiral of supermassive black-hole binaries and possible new early-universe sources.

Main Question

Can PTA data distinguish ULDM effects from ordinary SMBHB environments?

Pulsar timing arrays see a nanohertz gravitational-wave background, broadly consistent with supermassive black-hole binaries.

The low-frequency shape is where the physics enters:

- pure gravitational-wave driven binary evolution gives a benchmark spectrum.
- astrophysical environments can harden binaries (make their orbits shrink faster), suppressing low-frequency power.
- ultralight dark matter can also modify SMBHB evolution.

This talk: use predictive model comparison to ask what current PTA data can actually support.

Context: The Nanohertz Background Has Several Possible Origins

- Astrophysical baseline: an unresolved population of SMBHBs.
- Environmental hardening changes the residence time in the PTA band.
- New-physics alternatives include cosmic strings and dark-sector phase transitions.

Conference context

Other talks discuss nHz-background origins, including phase transitions and cosmic-string scenarios.

We assume an SMBHB origin and ask: **can ULDM be separated from generic environmental effects?**

Models Compared

ULDM models

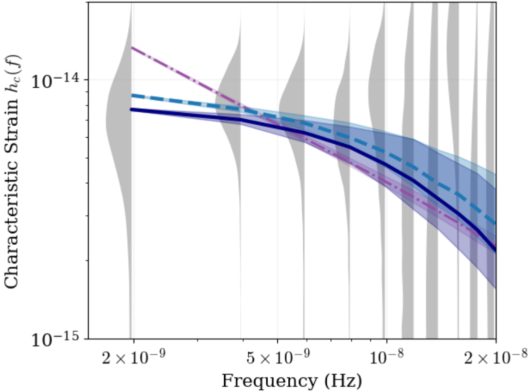
- **ULDM simplified**: assumes an idealized soliton profile around the SMBH binary.
- **ULDM realistic**: allows the binary to distort or pinch the soliton profile, changing the drag.

Benchmarks

- **Phenom**: NANOGrav benchmark using a double-power-law hardening timescale, with free time-scale and slope parameters that can produce a low-frequency turnover.
- **GW Only**: binaries evolve only through gravitational radiation.

All are tested on the five lowest PTA frequency bins used in the likelihood analysis.

Motivation: Why Modify GW-Only?



Background paper parameter-estimation fit: solid = realistic ULDM, dashed = simplified ULDM, magenta = GW only, grey violins = NANOGrav 15 yr.

- The lowest PTA frequencies are where environmental or ULDM-induced drag can suppress power.
- Treat this as motivation. The next step is to ask which models predict held-out bins better.

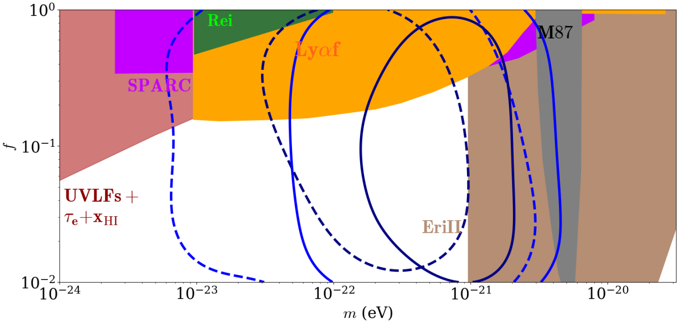
Previous Step: ULDM Constraints from PTA Data

ULDM picture

ULDM is a very light bosonic field with wave-like astrophysical structure. Around an SMBH binary, it can add drag, making the binary shrink faster and suppressing low-frequency PTA power.

Fit to NANOGrav 15 yr data:

- the preferred ULDM mass scale was around $10^{-21}\,\mathrm{eV}$.
- the inferred mass and effective fraction can lie in regions compatible with other constraints.
- this motivates a model-comparison question.



Background paper: solid/dashed contours show realistic/simplified ULDM fits over external constraints.

From Parameter Estimation to Model Comparison

Parameter estimation

If ULDM is the model, what particle masses and fractions are preferred or allowed?

Predictive model comparison

Do ULDM models predict held-out PTA frequency bins better than competing SMBHB models?

This distinction matters because a model can fit plausibly without being clearly preferred once predictive uncertainty is included.

Bayesian Leave-One-Out Cross-Validation (LOO-CV)

For each model and each frequency bin:

1. remove one PTA bin.
2. refit the model to the remaining bins.
3. evaluate the predictive density of the omitted bin.

Here LOO-CV means leaving out one PTA frequency bin at a time.

$$\widehat{\mathrm{elpd}}_{\mathrm{loo}} = \sum_{i=1}^{n} \log p(y_i \mid y_{-i})$$

elpd = expected log predictive density.

$$z_{\mathrm{loo}} = \frac{\Delta\widehat{\mathrm{elpd}}_{\mathrm{loo}}}{\Delta\widehat{\mathrm{SE}}}$$

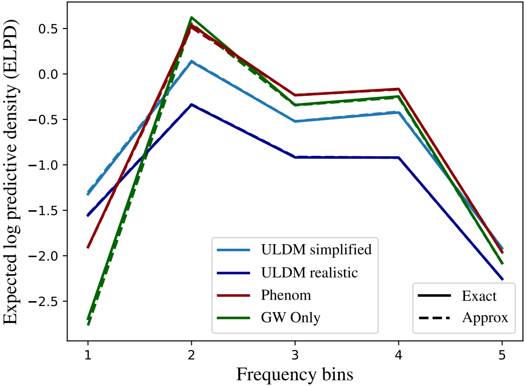
$$\Delta\widehat{\mathrm{SE}} = \sqrt{n\, s_d^2},$$

$$s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2.$$

d_i: elpd difference for bin i.

Important limitation: only a few low-frequency bins drive the separation, so the standard errors are useful diagnostics, not precision significance tests.

Where Do the Predictive Scores Come From?



Current paper: pointwise predictive contributions across the five PTA frequency bins.

- No single model dominates all frequency bins.
- GW Only is weakest in the lowest-frequency bin.
- The ULDM simplified model beats the ULDM realistic model bin-by-bin.

Exact LOO-CV Model Ranking

Model	$\widehat{\mathrm{elpd}}_{\mathrm{loo}}$	$\Delta\widehat{\mathrm{elpd}}$	z_{loo}
Phenom	−3.728	0.000	–
ULDM simplified	−4.053	0.324	0.36
GW Only	−4.741	1.012	1.34
ULDM realistic	−5.988	2.259	2.03

- The phenomenological environmental model has the largest total exact LOO-CV score.
- The separations from the best model are not large compared with the estimated errors.
- Current PTA data do **not** decisively prefer one model overall.

The Clearest Pairwise Pattern Is Within ULDM

Comparing ULDM simplified with ULDM realistic:

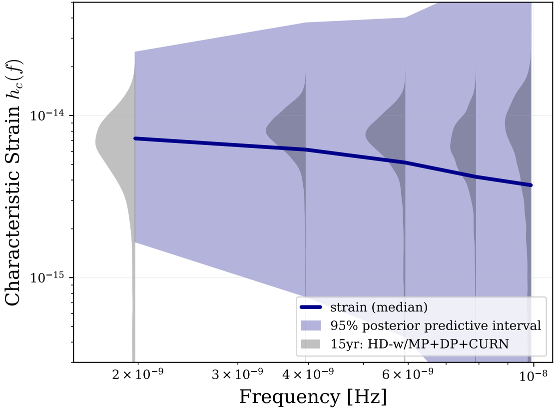
- simplified has larger pointwise predictive contribution in all five bins.
- $\Delta\widehat{\mathrm{elpd}}_{\mathrm{loo}} = 1.935$.
- $\Delta\widehat{\mathrm{SE}} = 0.238$.
- $z_{\mathrm{loo}} \approx 8$, but jackknife error $\sigma_{\mathrm{jack}} \approx 4$.

Interpretation

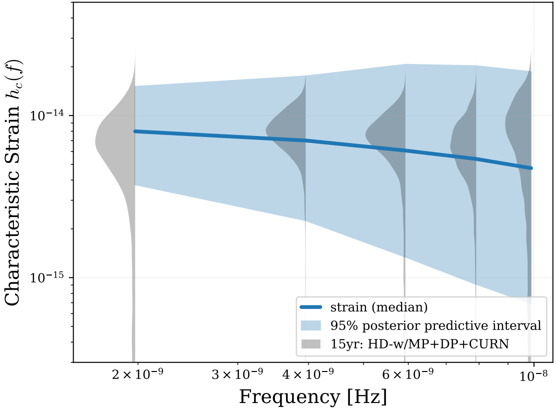
A large-looking standardized difference is still noisy because the comparison is driven by very few informative bins.

This is a descriptive pattern, not a statistically decisive preference for the simplified dynamics.

Why the ULDM Implementations Differ



ULDM realistic



ULDM simplified

The simplified predictive spectra are more concentrated near the observed low-frequency PTA data points.

Takeaways

1. PTA data remain compatible with ULDM-induced suppression of low-frequency power.
2. The earlier preferred mass/fraction region can be compatible with existing constraints.
3. In predictive comparison, a flexible environmental model ranks highest, but not decisively.
4. Current data do not yet separate ULDM from generic SMBHB environmental hardening.

What would help?

Better low-frequency PTA constraints, longer time baselines, and improved dynamical modelling of SMBHB environments.

Backup: ELPD Details

Leave-one-out predictive density:

$$p(y_i \mid y_{-i}) = \int p(y_i \mid \theta)\, p(\theta \mid y_{-i})\, d\theta.$$

Pointwise and total ELPD:

$$\mathrm{elpd}_{\mathrm{loo},i} = \log p(y_i \mid y_{-i}),$$

$$\mathrm{elpd}_{\mathrm{loo}} = \sum_{i=1}^{n} \mathrm{elpd}_{\mathrm{loo},i}.$$

Larger ELPD means better held-out predictive accuracy.

Monte Carlo estimate from exact refits:

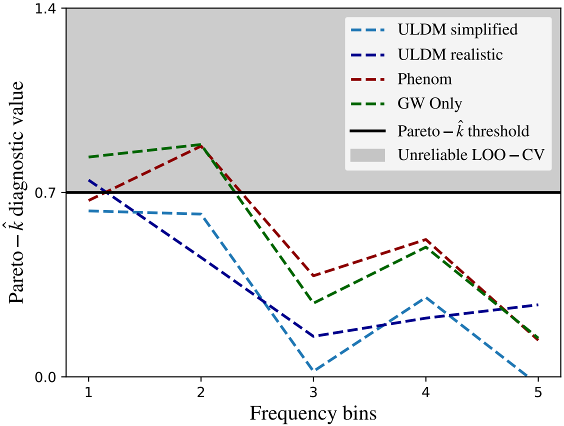
$$\theta_i^{(s)} \sim p(\theta \mid y_{-i}), \qquad s = 1, \ldots, S.$$

$$\widehat{\mathrm{elpd}}_{\mathrm{loo}} \approx \sum_{i=1}^{n} \log \left[\frac{1}{S} \sum_{s=1}^{S} p(y_i \mid \theta_i^{(s)}) \right].$$

Symbols

y_i: held-out PTA frequency bin. y_{-i}: all bins except i. θ: model parameters. $n = 5$: number of bins. S: number of posterior samples. $\theta_i^{(s)}$: sample s from the posterior fit without bin i.

Backup: PSIS Diagnostic



The PSIS approximation reuses the full-posterior chains instead of refitting the model.

The Pareto-$\hat{k}$ diagnostic checks whether the importance weights are reliable.

In this problem, some held-out frequency bins are influential, so the talk focuses on exact refits.

Backup: Phenom Model

Phenomenological environmental hardening:

$$\left.\frac{da}{dt}\right|_{\mathrm{phenom}} = H_a \left(\frac{a}{a_c}\right)^{1-\nu_{\mathrm{inner}}} \left(1+\frac{a}{a_c}\right)^{\nu_{\mathrm{inner}}-\nu_{\mathrm{outer}}} .$$

Total binary evolution rate:

$$\frac{da}{dt} = \left.\frac{da}{dt}\right|_{\mathrm{phenom}} + \left.\frac{da}{dt}\right|_{\mathrm{GW}} .$$

a is the binary semimajor axis. $a_c = 10^2\,\mathrm{pc}$ is the fixed break separation.

Hardening timescale:

$$t_h = \frac{dt}{d\ln a}.$$

Double-power-law limits:

$$\begin{aligned} t_h &\propto a^{\nu_{\mathrm{inner}}}, && a \ll a_c, \\ t_h &\propto a^{\nu_{\mathrm{outer}}}, && a \gg a_c. \end{aligned}$$

The target inspiral time, from the innermost stable circular orbit a_{ISCO} to a_{init}, fixes H_a:

$$\tau_f = \int_{a_{\mathrm{ISCO}}}^{a_{\mathrm{init}}} \frac{da}{|da/dt|}.$$

Varied parameters in the fiducial NANOGrav Phenom library

Population and black-hole–host parameters: $\psi_0, m_{\psi,0}, \mu, \epsilon_\mu$. Hardening-specific parameters: $\tau_f, \nu_{\mathrm{inner}}$. Fixed: $a_c = 10^2\,\mathrm{pc}$, $a_{\mathrm{init}} = 10^3\,\mathrm{pc}$, and $\nu_{\mathrm{outer}} = 2.5$. H_a is set by τ_f.