Quantifying Astrophysical Scattering With Transformers: Generic Mixture Density Network for FRB Analysis

Bikash Kharel[†], **Srinjoy Das**[‡], **Emmanuel Fonseca**[†]

[†]Department of Physics and Astronomy, West Virginia University
[‡]School of Mathematical and Data Sciences, West Virginia University
bk0055@mix.wvu.edu, {srinjoy.das, emmanuel.fonseca}@mail.wvu.edu

Abstract

The discovery rate of Fast Radio Bursts (FRBs) continues to increase with new radio facilities, yet extracting astrophysical parameters such as scattering timescale (τ) remains a significant bottleneck. Current approaches such as analytic template fitting and scattering-aware de-convolution are accurate but slow, initialization-sensitive, limited by low SNR, and require manual supervision. We present Multimodal Transformer Based Generic Mixture Density Network (**MT-GMDN**), which ingests FRB dynamic spectrum and timeseries through parallel transformer encoders, fuses their latent representations, and predicts the distribution of τ via a probabilistic mixture-density formulation that jointly handles the zero-inflated FRB population. Trained on $\sim$3,800 FRBs from CHIME/FRB Catalog 2, MT-GMDN achieves $R^2 = 94\%$ on scattering regression and 90% recall for detection, while running $\sim 10^4\times$ faster than classical fitting tools.

1 Introduction

Fast Radio Bursts (FRBs) are powerful extragalactic radio transients of millisecond duration (1). Propagation of the radiation through turbulent ionized interstellar and intergalactic media causes frequency-dependent temporal broadening modeled as a one-sided exponential tail with characteristic scattering timescale τ (2):

$$h(t) \propto \exp\left(-\frac{\Delta t}{\tau}\right), \qquad \tau \propto f^{-4}, \qquad I_{\rm obs}(t) = I_{\rm int}(t) * h(t). \tag{1}$$

Measuring τ has important astrophysical implications some of which include constraining plasma turbulence in host-galaxy environments (3; 4), diffuse halos (5), Milky Way ionized structure (6), and even cosmological parameters such as the Hubble constant (7). With detection rates reaching $\mathcal{O}(10^3\,\mathrm{FRBs\,yr^{-1}})$, scalable automated inference is now essential. Current tools such as `fitburst` (8) and Monte Carlo Markov Chain (MCMC) based estimation (9), are accurate but slow, initialization-sensitive, and can fail for events characterized by low SNR or high radio frequency interference (RFI). Deep learning in FRB science has focused on classification (10; 11) and detection (12), leaving parameter estimation largely unexplored. In this work we present **MT-GMDN**, which is the first Deep Learning based regression framework that jointly estimates τ, its heteroskedastic uncertainty, and the probability of unresolved scattering in a single forward pass.

Conference on Physics and AI at Stanford University (PAI 2026).

2 Methods

2.1 Probabilistic Formulation

Scattering timescales are strictly non-negative and empirically lognormally distributed (log-transforming τ on CHIME/FRB Catalog 2 (13) yields a near-Gaussian distribution). We model:

$$p(\tau \mid \mathbf{x}) = \mathrm{Lognormal}\big(\tau \mid \mu(\mathbf{x}),\, \sigma^2(\mathbf{x})\big)\,, \tag{2}$$

where location (μ) and scale (σ^2) parameters are estimated by a neural network. Since $>50\%$ of Catalog 2 bursts have unresolved scattering ($\tau < \tau_{\text{sampling time}}$), we extend Eq. 2 with a **Generic Mixture Density Network (GMDN)**:

$$p(\tau \mid \mathbf{x}) = p_0(\mathbf{x})\,\delta(\tau) + \big(1 - p_0(\mathbf{x})\big)\, f_{\mathrm{LN}}(\tau \mid \mathbf{x}), \tag{3}$$

where p_0 (Bernoulli parameter) is jointly learned with μ and σ (the parameters of the Lognormal f_{LN}) by minimizing cost function:

$$\ln p(\tau_i \mid \mathbf{x}_i) = \begin{cases} \ln p_{0,i}, & \tau_i = 0, \\ \ln(1 - p_{0,i}) + \ln \mathcal{N}(\ln \tau_i;\, \mu_i, \sigma_i^2) - \ln \tau_i, & \tau_i > 0. \end{cases} \tag{4}$$

This unifies detection and regression in one model, avoiding selection bias from discarding unresolved bursts. We also tested an alternative censored log-normal likelihood to handle unresolved cases reported as $\tau = 0$. Using the instrumental resolution limit, these samples were modeled as interval-constrained values ($0 \le \tau_i < 0.98\,\mathrm{ms}$) contributing $\log F_{\mathrm{LN}}(0.98 \mid \mu_i, \sigma_i)$ to the likelihood. However, this approach yielded highly unstable preliminary results, with predictions collapsing toward zero.

2.2 Model Architecture

The model architecture uses **two parallel transformer encoders** (Figure 1). The **Dynamic Spectrum Transformer** tokenizes two dimensional time frequency total intensity data $I(\nu, t)$ along time; each bin is a d-dimensional frequency embedding ($\mathbf{x}_{\mathrm{DS}} \in \mathbb{R}^{n \times d}$), with 4-head self-attention over two encoder layers capturing temporal dependencies and frequency-dependent smearing. The **Timeseries Transformer** operates identically on the 1-D spectral-mean pulse profile derived from $I(\nu, t)$, focusing on asymmetric tail decay. Learnable positional embeddings outperform sinusoidal variants (14). Latent outputs $\mathbf{z}_{\mathrm{DS}}, \mathbf{z}_{\mathrm{TS}} \in \mathbb{R}^{n \times d}$ from the transformers are concatenated, linearly projected, then aggregated by attention-based temporal pooling into $\mathbf{h} \in \mathbb{R}^d$. Three linear heads predict p_0 (sigmoid), μ (linear), and σ (softplus). **Ablation** studies confirm the importance of multimodal fusion. Single-branch Transformer models substantially degraded performance on both observational and simulated test sets: the dynamic-spectrum-only model generalized poorly with negative R^2, while the time-series-only model achieved only $\sim 72\%$ R^2. A CNN encoder with a mixture-density regression head also underperformed, reaching only $R^2 = 45\%$ on the observational test set. These results show that both modalities are essential for robust scattering-timescale estimation architecture.

2.3 Data and Preprocessing

Training uses CHIME/FRB Catalog 2 dynamic spectra: 16,384 frequency channels (400–800 MHz), $\sim$160 time samples at $0.983\,\mathrm{ms}$. 64-channel block averaging reduces to 256 channels; spectra are independently min-max normalized to $[0, 1]$. Target τ from `fitburst` are log-transformed ($\tau' = \ln \tau$ for $\tau > 0$, else 0). No SNR filtering is applied. Timeseries are generated by collapsing the dynamic spectrum along the frequency axis via spectral averaging, producing a 1-D pulse profile with n time samples. Because a single repeating FRB source can exhibit variable scattering timescales (15; 16) and burst morphologies (11), we randomly split the CHIME/FRB Catalog 2 into training, validation, and test sets. We also generated a separate simulated data set specifically for testing.

3 Results

3.1 Performance on CHIME/FRB Catalog 2

Regression ($R^2 = 94\%$): MT-GMDN predictions correlate strongly along $y = x$ against `fitburst` at $\nu_{\mathrm{ref}} = 400\,\mathrm{MHz}$. Constructed 95% CIs $e^{\mu \pm 1.96\sigma}$ capture the vast majority of catalogue values

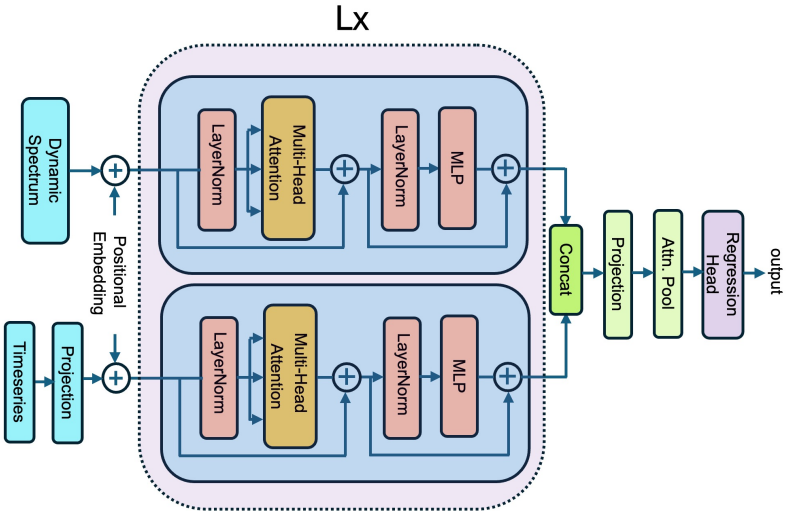


Figure 1: **MT-GMDN Architecture.** Dual-branch parallel Transformer simultaneously extracts features from dynamic spectra and timeseries simultaneously. Features are concatenated, compressed via attention-based pooling, and decoded by three heads parameterizing the zero-inflated log-normal likelihood via p_0, μ, and σ.

Table 1: MT-GMDN vs. MCMC on 50 simulated FRBs.

Metric	MT-GMDN (Lognormal)	MT-GMDN (Gamma)	MCMC	Target
RMSE (ms) $\downarrow$	7.62	**6.67**	7.83	Lower
$R^2 \uparrow$	0.71	**0.73**	0.69	Higher
Bias	-0.003	-0.98	-0.068	≈ 0
95% Coverage	1.00	**0.97**	0.89	≈ 0.95

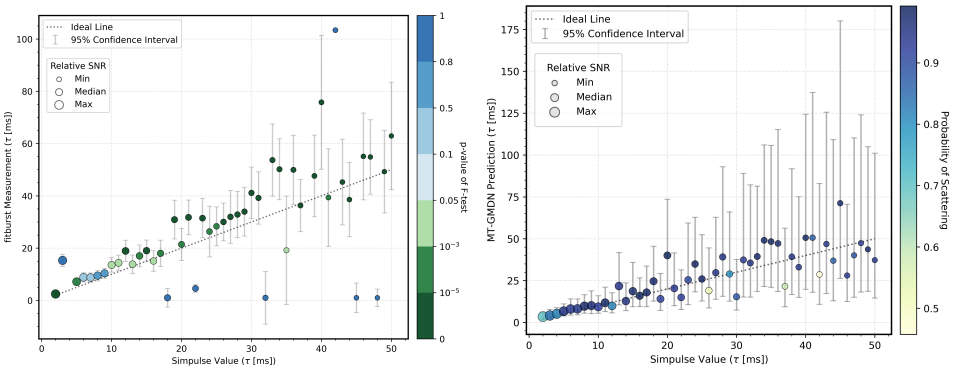
with uncertainty scaling correctly with τ. The conditional median e^{μ} serves as the point estimate for robustness to the heavy-tailed lognormal. **Scattering Detection (AUC** $= 0.87$**):** At threshold $p_0 = 0.40$: **90% recall**, 76% accuracy, 63% precision. False positives relative to the Catalog 2 are physically motivated as catalogue $\tau = 0$ entries are lower limits, not true zeros.

3.2 Evaluating MT-GMDN Against `fitburst` and MCMC on Simulated Data

Broadband synthetic FRBs at constant fluence have SNR decreasing with τ. `fitburst` yields $R^2 = 39\%$, with many non-significant F-test p-values used to assess scattering significance over the non-scattering model, and poorly calibrated CIs. In contrast, **MT-GMDN achieves** $R^2 = 68\%$, contains all targets within the 95% CI, and produces uncertainties that scale appropriately with SNR (Figure 2). After training, MT-GMDN has $\mathcal{O}(1)$ inference complexity and achieves $\sim 10^4\times$ speedup over `fitburst`. **Heteroskedasticity and Uncertainty Calibration:** Heteroskedastic behavior is confirmed by the Spearman correlation between predicted uncertainty σ and SNR: $\rho = -0.93$ ($p = 9.8 \times 10^{-18}$) for $\tau = 5, \mathrm{ms}$ and $\rho = -0.96$ ($p = 8.3 \times 10^{-23}$) for $\tau = 10, \mathrm{ms}$. To assess uncertainty calibration, we fit 50 simulated FRBs using MCMC with a scatter-broadened pulse model (17), treating its credible intervals as the reference. The Lognormal MT-GMDN shows over-coverage for high-scattering and/or low-SNR events due to its inflated tail, while replacing Lognormal components with Gamma distributions improves calibration and outperforms both MCMC and the Lognormal formulation. Results are summarized in Table 1 and Figure 3.

4 Discussion and Conclusion

MT-GMDN redefines FRB scattering inference by mapping burst morphology directly to τ, bypassing classical fitting bottlenecks. Multimodal fusion enables stable Transformer training with limited



(a) `fitburst` measurement (b) MT-GMDN measurement

Figure 2: Comparison on low-SNR synthetic pulses. Marker size encodes relative SNR; colour encodes `fitburst` F-test p-value (left) and MT-GMDN scattering probability (right). MT-GMDN maintains calibrated CIs where `fitburst` fails.

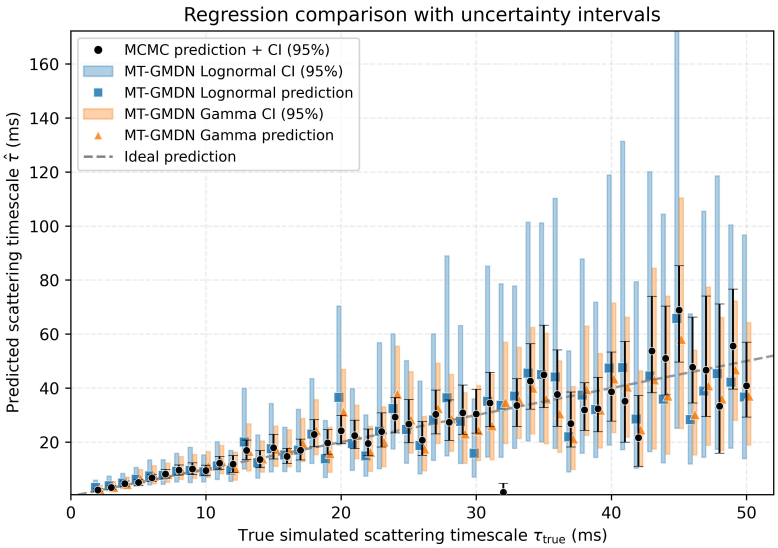


Figure 3: Comparison of predicted scattering timescales with 95% confidence intervals for MCMC, MTGMDN-Lognormal, and MTGMDN-Gamma, the SNR decreases from 20 to 3 going left to right.

data, while the GMDN framework handles zero-inflated targets through a principled probabilistic formulation. The model identifies subtle asymmetries near the instrumental resolution limit and signals buried in noise from heavy scattering, where traditional tools can fail, while operating at $\sim 10^4\times$ real-time speedup. However, modeling unresolved scattering with physically motivated likelihoods remains an open area for further research. Replacing the Lognormal mixture with a Gamma distribution mitigates heavy-tail over-coverage and yields better-calibrated confidence intervals, particularly for lower-SNR events. Beyond radio astronomy, the multimodal Transformer–GMDN framework offers a general template for probabilistic inference from low-SNR, limited-data time-frequency signals.

References

[1] Lorimer DR, Bailes M, McLaughlin MA, Narkevic DJ, Crawford F. A Bright Millisecond Radio Burst of Extragalactic Origin. *Science*. 2007 Nov;318(5851):777.

Code available at `https://github.com/kharelb/Scattering-Timescale-Estimation-With-Transformers`

[2] Williamson IP. Pulse broadening due to multiple scattering in the interstellar medium. III. . 1974 Mar;166:499-512.

[3] Xu S, Zhang B. On the Origin of the Scatter Broadening of Fast Radio Burst Pulses and Astrophysical Implications. . 2016 Dec;832(2):199.

[4] Shin K, Leung C, Simha S, Andersen BC, Fonseca E, Nimmo K, et al. Investigating the sightline of a highly scattered Fast radio burst through a cosmic sheet structure in the local universe. Astrophys J. 2025 Nov;993(2):208.

[5] Ocker SK, Cordes JM, Chatterjee S. Constraining Galaxy Halos from the Dispersion and Scattering of Fast Radio Bursts and Pulsars. . 2021 Apr;911(2):102.

[6] Shivraj Patil S, Main RA, Fonseca E, McGregor K, Gaensler BM, Brar C, et al. A Spatial Gap in the Sky Distribution of Fast Radio Burst Detections Coinciding with Galactic Plasma Overdensities. arXiv e-prints. 2025 Sep:arXiv:2509.06721.

[7] Yang, Tsung-Ching, Hashimoto, Tetsuya, Hsu, Tzu-Yin, Goto, Tomotsugu, Ling, Chih-Teng, Ho, Simon C -C , et al. Constraining the Hubble constant with scattering in host galaxies of fast radio bursts. AA. 2025;693:A85. Available from: https://doi.org/10.1051/0004-6361/202450823.

[8] Fonseca E, Pleunis Z, Andersen BC, Bhardwaj M, Boyle PJ, Chawla P, et al. Modeling the Morphology of Fast Radio Bursts and Radio Pulsars with fitburst. The Astrophysical Journal Supplement Series. 2024 Mar;272(1):7.

[9] CHIME/FRB Collaboration, Amiri M, Bandura K, Bhardwaj M, Boubel P, Boyce MM, et al. Observations of fast radio bursts at frequencies down to 400 megahertz. . 2019 Jan;566(7743):230-4.

[10] Connor L, van Leeuwen J. Applying Deep Learning to Fast Radio Burst Classification. The Astronomical Journal. 2018 nov;156(6):256. Available from: https://doi.org/10.3847/1538-3881/aae649.

[11] Kharel B, Fonseca E, Brar C, Khan A, Mas-Ribas L, Patil SS, et al.. Repeating vs. Non-Repeating FRBs: A Deep Learning Approach To Morphological Characterization; 2025. Available from: https://arxiv.org/abs/2509.06208.

[12] Agarwal D, Aggarwal K, Burke-Spolaor S, Lorimer DR, Garver-Daniels N. FETCH: A deep-learning based classifier for fast transient classification. Monthly Notices of the Royal Astronomical Society. 2020 Sep;497(2):1661-74.

[13] Chime/Frb Collaboration, Abbott T, Andersen BC, Andrew S, Bandura K, Bhardwaj M, et al. The Second CHIME/FRB Catalog of Fast Radio Bursts. . 2026 Mar;283(1):34.

[14] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al.. Attention Is All You Need; 2023. Available from: https://arxiv.org/abs/1706.03762.

[15] Pandhi A, Nimmo K, Andrew S, Brar C, Chatterjee S, Cook AM, et al. A Steadily Declining Dispersion Measure for the Repeating Fast Radio Burst FRB 20220529A: Evidence for a Fast Radio Burst Engine Embedded in an Expanding Supernova Remnant. . 2026 Apr;1000(2):L53.

[16] Ocker SK, Cordes JM, Chatterjee S, Li D, Niu CH, McKee JW, et al. Scattering variability detected from the circumsource medium of FRB 20190520B. . 2023 Feb;519(1):821-30.

[17] McKinnon MM. The Analytical Solution to the Temporal Broadening of a Gaussian-Shaped Radio Pulse by Multipath Scattering from a Thin Screen in the Interstellar Medium. Publications of the Astronomical Society of the Pacific. 2014 may;126(939):476. Available from: https://doi.org/10.1086/676975.