Using Multi-Modal Diffusion Models to Reconstruct Dark Matter Fields

Aditi Chandrashekar *, **Saumya Chauhan** *, **Eshani Patel** *, **Maria Vazhaeparambil** *

[1]California Institute of Technology,
1200 E California Blvd, Pasadena, CA 91125
{aditijc, saumya.s.chau, i.eshanipatel, maria.j.vazhaeparambil}@gmail.com

Abstract

The cosmic web, governed by dark matter, shapes the universe's large-scale structure. Reconstructing these dark matter maps is challenging because galaxies are inherently biased tracers of the underlying distribution. Diffusion models have recently shown strong promise in mitigating such biases, but existing approaches are limited by their reliance on a single tracer—typically stellar mass—even though real observational data are rich and inherently multimodal. We introduce a multimodal diffusion framework conditioned on stellar mass, fast radio burst (FRB) dispersion measures, and gravitational lensing, three complementary tracers that together offer a more complete view of matter distribution. Applied to simulations from the CAMELS suite, this framework yields high-fidelity reconstructions of dark matter fields. Crucially, our approach goes beyond reconstruction: by systematically varying input signal-to-noise ratios (SNR), it gives insight into the mapping between instrument noise levels and expected reconstruction gains. This enables principled, modality-aware survey design and instrument planning, identifying where improvements in sensitivity have the highest scientific payoff. Code to reproduce these experiments can be found at https://github.com/epatel16/Dark-Matter-Diffusion.git.

Introduction

The cosmic web is predominantly shaped by dark matter, a quantity that cannot be observed directly, but whose gravitational influence governs galaxy structure formation. Galaxies offer only biased tracers of this underlying web, as their distributions are influenced by uncertain astrophysical processes. Accurately reconstructing dark matter fields is therefore crucial to disentangle these effects and improve cosmological inference. Recent work has demonstrated that diffusion generative models can reconstruct unbiased dark matter fields from stellar mass maps within the CAMELS simulation suite (Ono et al. 2024). We extend this framework by conditioning on additional complementary modalities—fast radio bursts (FRBs) and gravitational lensing—to exploit their distinct tracers of baryonic and total mass distributions (Ravi 2019; Bartelmann and Schneider 2001).

Alternative deep generative approaches for cosmological fields include variational autoencoders, normalizing flows, and generative adversarial networks (Kingma and Welling 2014; Rezende and Mohamed 2015; Goodfellow et al. 2014). In our setting, however, two properties of diffusion models are particularly important. First, a DDPM is explicitly trained to denoise fields at a continuum of noise levels (Ho, Jain, and Abbeel 2020; Kingma et al. 2021; Song et al. 2020). We can align these diffusion time steps with realistic instrument noise levels for each tracer, which allows a single trained model to forecast reconstruction quality across a grid of stellar, FRB, and lensing SNRs without retraining. Second, conditional diffusion treats each tracer as an input channel in a shared spatial field, so the network can learn cross–modal interactions while still allowing us, at test time, to "turn off" a tracer or change its noise level and observe how the inferred dark–matter map responds. These aspects make diffusion models especially well matched to our goal of multi–tracer reconstruction and SNR–aware survey design.

Recent work, such as Park et al. (2024), has demonstrated the power of diffusion models for volumetric dark matter reconstruction using galaxy survey data. Our study complements these 3D approaches: rather than focusing on volumetric inference, we develop a flexible 2D framework that integrates stellar, FRB, and lensing inputs in a controlled setting. This design enables systematic analysis of the contribution and robustness of different tracers under varying noise levels, providing a foundation that can be extended or incorporated into future large-scale 3D reconstruction pipelines.

Our contributions include:

1. A multimodal diffusion model that jointly leverages three complementary modalities to reconstruct dark matter maps.
2. A framework using this model to forecast reconstruction quality given future measurement signal-to-noise ratios (SNRs).

Given fixed resources we address the question of which modality benefits most from marginal SNR improvements, and speculate as when this investment may no longer benefit reconstruction. Our framework provides empirical results to substantiate this for stellar, FRB, and lensing inputs.

Methods

Conditional Diffusion Model. We implement a multimodal denoising diffusion probabilistic model (DDPM)

*These authors contributed equally.

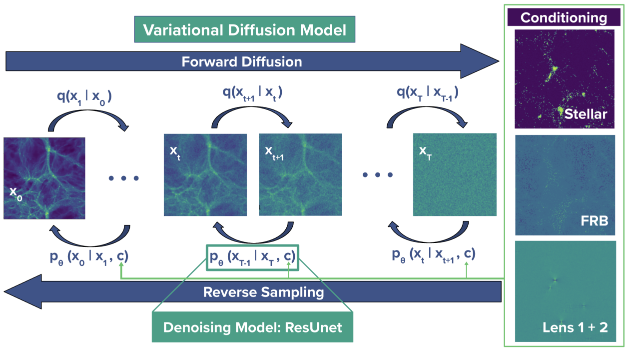

Figure 1: We propose a multimodal variational diffusion model for reconstructing dark matter maps from measurement data, conditioning on stellar, FRB, and lensing inputs.

with residual and attention layers, expanding upon the unimodal model discussed in Ono et al. (2024). At each timestep, stellar mass, FRB, and lensing maps provide spatial guidance for denoising. We train with a cosine annealing learning rate schedule (Ho, Jain, and Abbeel 2020). Models were trained for 30 epochs with 1000 diffusion steps, a learning rate of 1×10^{-4}, batch size 12, and 256×256 resolution images. The baseline multimodal model is configured with stellar noise $\sigma = 0.1$, lensing noise $\sigma = 10$, and fast radio burst noise $\sigma = 10$. For each single-modality model, we adopt the corresponding noise level from the multimodal setting.

Datasets and Preprocessing. We use 2D maps from the CAMELS Astrid simulation at $z = 0$ redshift (Villaescusa-Navarro et al. 2021). Stellar mass, FRBs, and gravitational lensing trace the cosmic web through stellar populations, ionized gas dispersion (Ravi 2019), and light deflection (Bartelmann and Schneider 2001; Schneider, Ehlers, and Falco 2006). Dark matter, stellar, and gas maps are log-transformed and globally normalized. Stellar mass is estimated via Stellar Population Synthesis and provided as 2D maps in the CAMELS simulations, and is tightly correlated with dark matter, serving as a direct tracer of the cosmic web. Fast Radio Bursts (FRBs) probe the ionized gas distribution through their dispersion measures, offering an indirect measurement of dark matter as gas follows its gravitational potential. Gravitational lensing directly traces dark matter by measuring how mass bends light, providing a powerful complement to galaxy-based tracers. FRBs are simulated from gas maps by randomly masking pixels, where the retained fraction defines the effective SNR. Lensing shear fields g_1 and g_2 are generated from mass maps by first computing convergence κ, transforming to Fourier space, applying lensing kernels, and inverse FFT.

Evaluation Metrics. We assess reconstruction fidelity using four complementary metrics. The **power spectrum** $P(k)$ measures the variance of matter density fluctuations as a function of scale (wavenumber k), quantifying clustering strength across spatial scales. For a density fluctuation field $\delta(x) = \frac{\rho(x) - \bar{\rho}}{\bar{\rho}}$ the Fourier transform $\delta(k)$ defines the power spectrum as $P(k) = \langle |\delta(k)|^2 \rangle$, where $\langle \cdot \rangle$ denotes averaging over modes with the same wavenumber k. Comparing $P(k)$ between reconstructions and ground truth tests whether the correct distribution of structure across scales is recovered. The **cross-correlation coefficient** $R(k)$ measures phase alignment between reconstructed and true fields, defined as $R(k) = \frac{P_{\text{true,sample}}(k)}{\sqrt{P_{\text{true}}(k)}\sqrt{P_{\text{sample}}(k)}}$, with values near 1 indicating strong agreement across scales. We additionally compute the peak signal-to-noise ratio (PSNR), which evaluates pixel-level fidelity relative to noise, and the mean squared error (MSE), which measures the average squared difference between generated and true maps. For robustness, we generate 10 samples per guide and report the mean and variance across SNR levels.

Results

Multimodal Model. We compare our multimodal model to the baseline unimodal models corresponding to the stellar, fast radio burst, and gravitational lensing modalities and quantify the reconstruction quality through perceptual metrics, power spectra, and cross-correlation evaluation (figure 2 and Supplementary Materials). Stellar data with low noise is the most effective single modality for reconstructing dark matter fields. Lensing data provides weak signal, while FRBs offer moderate signal. Our multimodal model performs best, achieving more consistent correlations.

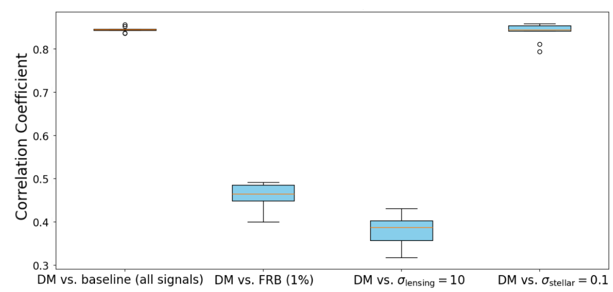

Figure 2: Dark matter cross correlation coefficients (exponentiated) from our multimodal model vs. three single modality models, computed over 10 random test samples.

Forecasting Future SNR (Stress Test). We stress test the model by evaluating it under tracer SNR conditions that differ from the baseline settings used during training. Concretely, we systematically vary the noise standard deviation σ for each modality (stellar, FRB, lensing) while keeping the network weights fixed, and quantify how reconstruction fidelity changes. Fidelity is measured by power spectrum similarity and cross-correlation with true dark-matter fields, as well as PSNR and MSE (Table 1 and Figure 3). As σ decreases, the SNR increases, so this procedure probes both degraded and improved measurement regimes along the diffusion noise axis without retraining the model. This out-of-distribution SNR sweep reveals where the learned conditional mapping remains robust and where reconstruction

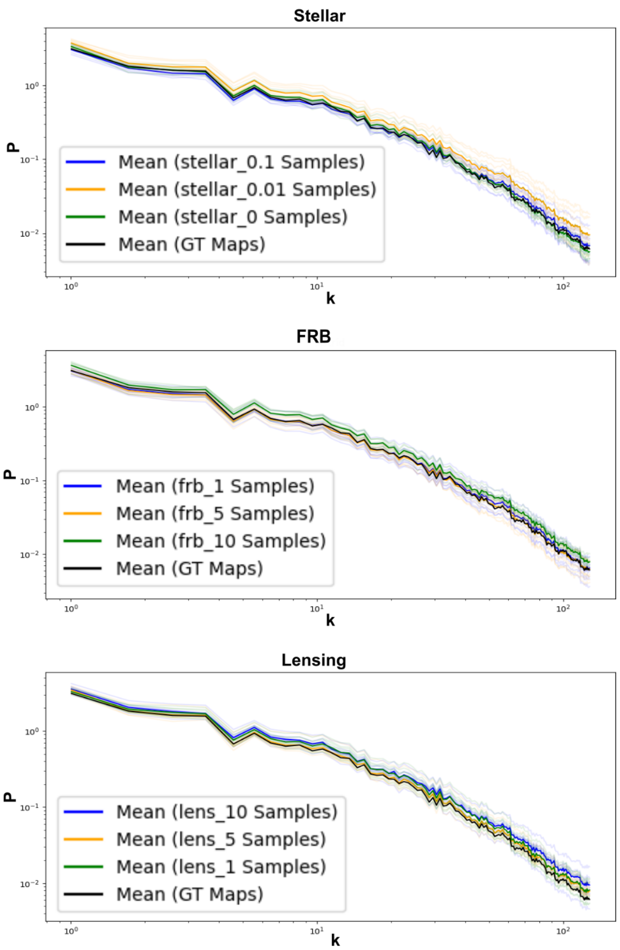

Figure 3: Forecasting future SNR values. Black line is power spectra of the ground truth dark matter maps.

quality begins to break down, providing the basis for our survey-design forecasts.

Stellar σ	Lensing σ	FRB σ	PSNR	MSE
0.1	10	1	30.798	0.187
0.01	10	1	30.653	0.192
0	10	1	28.385	0.320
0.1	10	1	30.798	0.187
0.1	5	1	29.775	0.240
0.1	1	1	28.232	0.322
0.1	10	10	29.637	0.248
0.1	10	5	29.850	0.231
0.1	10	1	30.798	0.187

Table 1: Model performance when varying noise standard deviation (σ) for stellar, lensing, and FRB maps. Rows are grouped by the varied modality.

Figures 4 and 5 illustrate additional results from our experiments varying signal-to-noise ratios (SNR) across stellar, FRB, and lensing modalities.

The cross-correlation plots (Fig. 4) show that higher SNR consistently improves alignment between reconstructed and ground-truth dark matter maps. For stellar and FRB inputs, correlation increases sharply as SNR improves, highlighting their sensitivity to noise. Lensing inputs show a more gradual trend, suggesting greater robustness to degraded SNR.

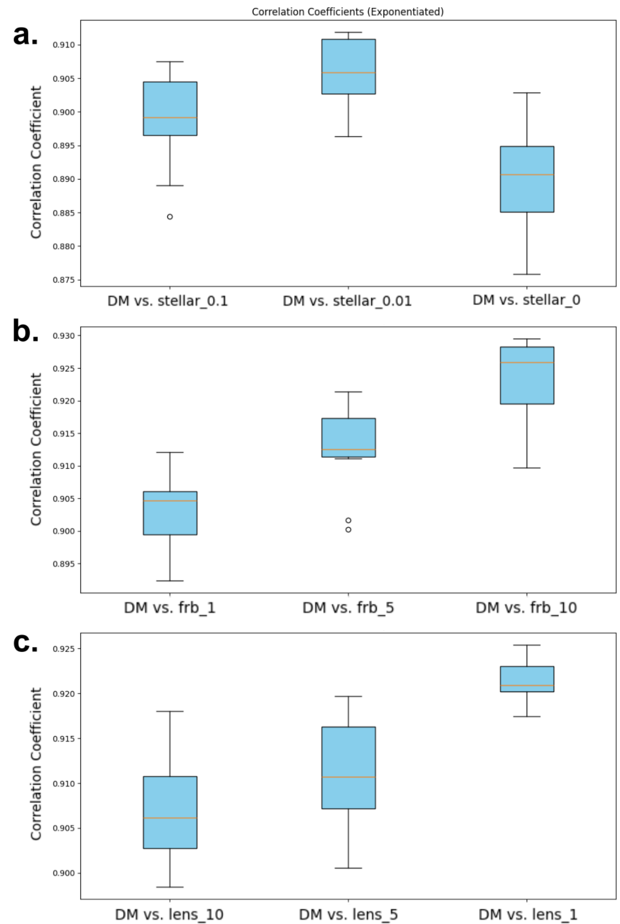

Figure 4: Dark Matter map cross correlation plots for experiments with varied stellar, frb, and lensing SNRs respectively. Note that these values are exponentiated. Metrics are averaged over 10 generated samples.

The power spectrum correlation plots (Fig. 5) reinforce this trend. At high SNR, correlations remain stable across spatial scales, while at low SNR the variance increases substantially, particularly for stellar inputs. FRB correlations remain noisier overall, while lensing inputs achieve relatively stable performance even at moderate SNR.

Takeaways: We find that (i) Stellar inputs are the most informative but also the most sensitive to noise. (ii) FRB inputs benefit from higher SNR but remain less stable. (iii) Lensing inputs are comparatively robust, though they carry less discriminative power at fine scales. These results emphasize the complementary roles of modalities: stellar data

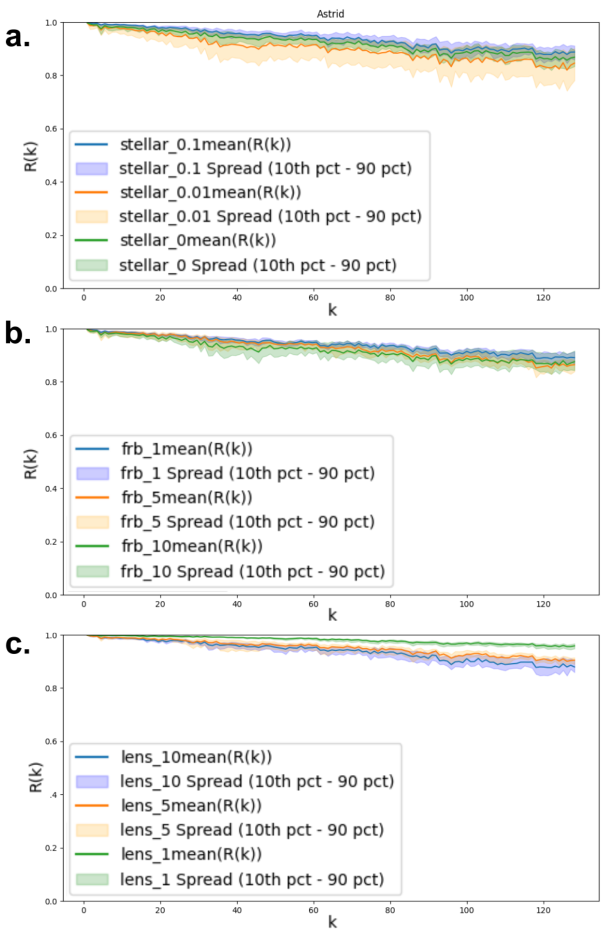

Figure 5: Power spectra correlation plots for experiments with varied stellar, frb, and lensing SNRs respectively. Metrics are averaged over 10 generated samples.]

excels when clean, lensing contributes stability under noisy conditions, and FRBs add value primarily when SNR is sufficiently high.

Discussion & Conclusion

We introduced a multimodal diffusion framework that conditions on three complementary tracers—stellar mass, FRB dispersion measures, and weak-lensing shear—to reconstruct dark-matter fields in the CAMELS Astrid suite. Across $P(k)$, scale-wise $R(k)$, and PSNR/MSE, the multimodal model outperforms single-modality baselines, indicating that each tracer contributes non-redundant information. Compared with alternative deep generative architectures such as VAEs, flows, or GANs (Kingma and Welling 2014; Rezende and Mohamed 2015; Goodfellow et al. 2014), our conditional diffusion model is explicitly organized around a noise axis and compositional conditioning on multiple tracers, which we exploit to convert tracer SNR into quantitative, uncertainty-aware forecasts for survey and instrument design. Beyond reconstruction accuracy, a central contribution is a *design-facing* analysis that converts per-modality SNR into expected reconstruction quality, providing quantitative guidance for future survey and instrument choices.

Experiment-design takeaways. By sweeping input noise σ for each modality, we obtain SNR–performance curves that (i) identify *high-leverage regimes* where marginal SNR gains translate to large improvements in $P(k)$ and $R(k)$, and (ii) expose *saturation regimes* where further sensitivity yields diminishing returns. Empirically, we find steep initial gains for stellar maps that flatten at moderate SNR, threshold-like improvements for FRB once past a noise floor, and monotonic, stable gains for lensing that are broad across scales. These modality-specific slopes and breakpoints enable principled budget allocation—e.g., prioritizing FRB/lensing upgrades once stellar sensitivity enters a plateau.

Extensibility of uncertainty-aware forecasting. Because the diffusion model yields sample ensembles, our forecasts report means and variances of metrics over repeated generations at fixed SNR. This provides uncertainty bands on the SNR–performance curves, which are important when comparing design options with similar expected means but different risk profiles across spatial scales. The framework is modular: additional tracers (e.g., HI intensity maps, SZ/thermal dust surrogates) can be incorporated with minimal architectural changes, and instrument-specific noise models can directly re-parameterize our SNR axis. In practice, one can (1) plug in a candidate survey's forward noise model, (2) read off expected $P(k)/R(k)$ improvements at target SNRs, and (3) choose operating points that maximize scientific return per marginal sensitivity dollar. Our study is complementary to volumetric inference approaches that operate directly in 3D with galaxy survey data (e.g., Park et al. (2024)). The proposed 2D multimodal formulation emphasizes modality fusion, controllability, and design-oriented SNR exploration; it can serve upstream of 3D reconstructions as a prior/initializer or as a tool to set tracer-specific sensitivity targets before full volumetric modeling.

Limitations and outlook. Results are shown at $z=0$ on simulated maps and do not yet include domain gaps (selection functions, PSFs, masks) inherent to observations. Future work will (i) incorporate realistic survey systematics and mask-aware conditioning, (ii) add tracers with distinct systematics to refine trade-off analyses, and (iii) couple the forecasting module to cosmological parameter inference, yielding an end-to-end, observation-ready pipeline where SNR design targets are optimized for downstream parameter constraints, not just field-level fidelity.

Conclusion. In summary, the proposed multimodal diffusion model improves dark-matter field reconstruction and, crucially, turns instrument SNR into *decision-relevant* forecasts. These forecasts reveal where additional sensitivity pays off, where it saturates, and how tracer combinations can be balanced to meet science goals—providing a practical bridge between generative reconstruction and experimental design.

References

Bartelmann, M.; and Schneider, P. 2001. Weak gravitational lensing. *Physics Reports*, 340(4): 291–472.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*, volume 27.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.

Kingma, D.; Salimans, T.; Poole, B.; and Ho, J. 2021. Variational diffusion models. *Advances in neural information processing systems*, 34: 21696–21707.

Kingma, D. P.; and Welling, M. 2014. Auto-Encoding Variational Bayes. *Proceedings of the 2nd International Conference on Learning Representations (ICLR)*. ArXiv:1312.6114.

Ono, V.; Park, C. F.; Mudur, N.; Ni, Y.; Cuesta-Lazaro, C.; and Villaescusa-Navarro, F. 2024. Debiasing with Diffusion: Probabilistic reconstruction of Dark Matter fields from galaxies with CAMELS. *arXiv preprint arXiv:2403.10648*.

Park, C. F.; Mudur, N.; Cuesta-Lazaro, C.; Ni, Y.; Ono, V.; and Finkbeiner, D. 2024. 3D Reconstruction of Dark Matter Fields with Diffusion Models: Towards Application to Galaxy Surveys. In *ICML 2024 AI for Science Workshop*.

Ravi, V. 2019. Measuring the circumgalactic and intergalactic Baryon contents with fast radio bursts. *The Astrophysical Journal*, 872(1): 88.

Rezende, D. J.; and Mohamed, S. 2015. Variational Inference with Normalizing Flows. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*.

Schneider, P.; Ehlers, J.; and Falco, E. E. 2006. *Gravitational Lenses*. Springer.

Song, Y.; Sohl-Dickstein, J.; Kingma, D. P.; Kumar, A.; Ermon, S.; and Poole, B. 2020. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*.

Villaescusa-Navarro, F.; et al. 2021. The CAMELS project: Cosmology and astrophysics with machine-learning simulations. *The Astrophysical Journal*, 915(1): 71.