
Dynamic Diffusion Schrödinger Bridge in Astrophysical Observational Inversions

Ye Zhu^{*1,2}, Duo Xu³, Zhiwei Deng⁴, Jonathan C. Tan^{5,6}, Olga Russakovsky¹

¹Department of Computer Science, Princeton University, Princeton NJ, USA

²LIX, École Polytechnique, IP Paris, Palaiseau, France

³Canadian Institute for Theoretical Astrophysics (CITA), University of Toronto, Toronto, Canada

⁴Google DeepMind, Mountain View, CA, USA

⁵Department of Astronomy, University of Virginia, Charlottesville VA, USA

⁶Department of Space, Earth & Environment, Chalmers University of Technology, Gothenburg, Sweden

ye.zhu@polytechnique.edu, xuduo@cita.utoronto.ca,
zhiweideng@google.com, jct6e@virginia.edu, olgarus@princeton.edu

Abstract

We study Diffusion Schrödinger Bridge (DSB) models in the context of dynamical astrophysical systems, specifically tackling observational inverse prediction tasks within Giant Molecular Clouds (GMCs) for star formation. We introduce the *Astro-DSB* model, a variant of DSB with the pairwise domain assumption tailored for astrophysical dynamics. By investigating its learning process and prediction performance in both physically simulated data and in real observations (the Taurus B213 data), we present two main takeaways. First, from the astrophysical perspective, our proposed paired DSB method improves interpretability, learning efficiency, and prediction performance over conventional astrostatistical and other machine learning methods. Second, from the generative modeling perspective, probabilistic generative modeling reveals improvements over discriminative pixel-to-pixel modeling in Out-Of-Distribution (OOD) testing cases of physical simulations with *unseen initial conditions and different dominant physical processes*. Our study expands research into diffusion models beyond the traditional visual synthesis application and provides evidence of the models' learning abilities beyond pure data statistics, paving a path for future physics-aware generative models which can align dynamics between machine learning and real (astro)physical systems.²

1 Introduction

Diffusion models (DMs) (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song et al., 2020), sometimes also known as score-based generation models (SGMs), have become a popular probabilistic modeling approach in machine learning for many data generation applications (Dhariwal and Nichol, 2021; Ho et al., 2022; Rombach et al., 2022). A recent line of work improves the vanilla design by reformulating the mapping trajectory learning in the context of the Schrödinger bridge problem (Schrödinger, 1932), which allows for imposing domain-specific restrictions on the learned trajectory, leading to Diffusion Schrödinger Bridge models (DSB) (Wang et al., 2021; De Bortoli et al., 2021; Shi et al., 2023). These DSB models are theoretically grounded and have proven effective in vision-related applications such as image generation and translation (Liu et al., 2023; Gushchin et al., 2024). However, their modeling capabilities and limits could extend beyond vision, offering a new opportunity to test these generative

^{*}Work mainly completed when YZ was a postdoc at Princeton University.

²Code available at <https://github.com/L-YeZhu/AstroDSB>

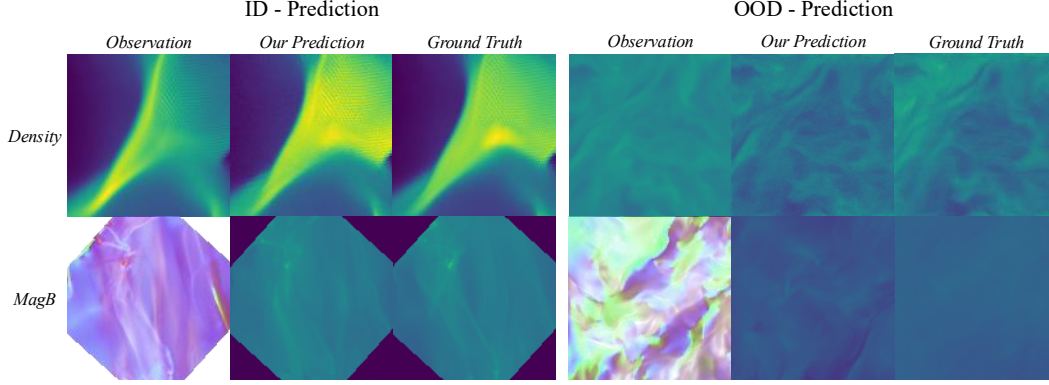


Figure 1: **We introduce *Astro-DSB* for astrophysical inverse predictions from observables within Giant Molecular Clouds (GMCs).** We visualize the observables, predicted results, and ground truth physical distributions on the density (top) and magnetic field (bottom) for in-distribution (*ID*) and out-of-distribution (*OOD*) testing cases, with the *OOD* defined via changes in initial physical conditions and dominant dynamical processes in the simulations.

formulations in the context of real-world physical systems, where data are governed by complex dynamics and observational constraints rather than pixel statistics.

In this work, we study a real-world physical problem, focusing on a key class of inverse prediction tasks in astrophysics, particularly within the context of star formation (McKee and Ostriker, 2007; Larson, 2003; Madau and Dickinson, 2014). Specifically, we aim to better understand and predict the *physical states* within Giant Molecular Clouds (GMCs) (Shu et al., 1987), which is a major component of the interstellar medium (ISM) (Spitzer Jr, 2008) that directly regulates star formation rates. The goal of our task is to infer true physical distributions, such as density or magnetic fields, from partial observables, as illustrated in Fig. 1.

As a critical and active research area in astrophysics, multiple methods have been developed over the past decades to tackle this challenge. Given that observational inverse prediction problems are inherently intractable, early approaches often relied on *astrostatistical methods* such as power-law fitting (Maschberger and Kroupa, 2009), which combine theoretical simplifications with observational statistical priors (detailed in Sec. 2.2 and Sec. 4.2). However, these methods typically suffer from large prediction errors and limited generalization on out-of-distribution (*OOD*) testing cases under shifted physical initial conditions, due to often oversimplified physical assumptions and restricted capacity for modeling complex non-linearities.

Given the increasing capability of neural networks in modeling large-scale data with highly complicated nonlinearities, the astrophysical community has recently started to explore machine learning approaches (Yang et al., 2024; Xu et al., 2020a) to overcome previous limitations. Among multiple ML architectures, conditional diffusion generative models (DGMs) (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song et al., 2020) stand out as recent state-of-the-art (SOTA) frameworks with large improvements in prediction accuracy over conventional methods for several critical physical quantities, such as GMCs density distribution (Xu et al., 2023b) and magnetic field strengths (Xu et al., 2025). These methods typically take the observables as input conditions and output the predicted physical distribution maps through the distribution mapping ability of DGMs trained on simulation data. However, despite the enhanced prediction accuracy, such conditional diffusion models share a few constraints, including a physics-ML misalignment due to the Gaussian prior from the original DM formulation (also detailed in Sec. 3.2), as well as the known expensive computational cost for learning DGMs.

To this end, we introduce *Astro-DSB*, a variant of Diffusion Schrödinger Bridge model tailored to astrophysics with better physics-ML state alignment and improved learning efficiency. Specifically, compared to existing DGMs-based methods that learn a mapping between the Gaussian prior and the true physical distribution with the observable as the condition, *Astro-DSB* removes the reliance on the Gaussian prior from the DGMs formulation and learn the direct mapping between observables and ground-truth physical states. This design also helps to *accelerate the learning efficiency by 75%*

compared to the previous DGM-based SOTA. More concretely, *Astro-DSB* leverages the intrinsic deterministic mapping property between the observation and ground truth physical state and adopts a paired assumption between two boundary distributions³. Within this high-level Schrödinger bridge ML framework, we further integrate the observables as an enhanced condition to facilitate the bridge learning process. Intuitively, this can be mathematically considered as learning a *conditional Doob’s h-transform*, thus differentiating from the vanilla paired DSB models in computer vision tasks (Liu et al., 2023) and scientific applications such as protein designs (Didi et al., 2023).

Our comprehensive experiments on both molecular density and magnetic field tasks validate the effectiveness of our proposed method. In addition to the comparable or superior SOTA performance in ID testing cases, we also examine the performance under distribution shifts using both physically simulated data and **real astrophysical observations from Taurus B213**. In contrast to conventional vision tasks, where OOD is often heuristically defined, astrophysical systems provide scientifically meaningful definitions of distributional shifts through variations in initial conditions and dominant physical processes (Wang and Abel, 2009; Wu et al., 2015). These configurations reflect realistic physical diversity, making them especially suitable for testing distribution-level generalization. Through this evaluation, we uncover the practical boundaries of probabilistic modeling and its advantages in scientific domains such as astrophysics.

The rest of the paper is organized as follows: we present related works in Sec. 2; Sec. 3 details the core technical designs; Sec. 4 explains our controlled experiments and comprehensive analysis comparing against conventional astrophysical and existing ML approaches; and we conclude by discussing the limitations and future directions in Sec. 5. To summarize our main contributions, our work offers two key takeaway messages: for scientific applications such as astrophysical inverse prediction, the proposed *Astro-DSB* framework offers an efficient and accurate alternative to conventional modeling approaches; for the broader generative modeling community, our study highlights the advantage of distribution-level learning in enabling meaningful generalization under physically grounded distribution shifts, extending the application scenarios from simulation data to real-world observations from Taurus B213 (Li and Goldsmith, 2012; Palmeirim et al., 2013).

2 Related work

2.1 Schrödinger bridge problem and generative modeling

The machine learning method we leverage in this work is closely related to one line of active research efforts in the Schrödinger bridge problem (SB) with application to generative models. Specifically, given the similar underlying problem of learning the mapping between two arbitrary probabilistic distributions, prior works (Wang et al., 2021; De Bortoli et al., 2021; Shi et al., 2023, 2022; Somnath et al., 2023; Liu et al., 2023) have explored improving generative modeling by formulating the boundary distribution matching problem within the Optimal Transport (OT) map of SB (Cuturi, 2013). Specifically, Wang et al. (2021) proposes to learn a generative model via entropy interpolation with a Schrödinger bridge in a two-stage manner. While this work relies on a *static* SB problem, later works (De Bortoli et al., 2021; Shi et al., 2023; Liu et al., 2023; Gushchin et al., 2024; Ksenofontov and Korotin, 2025) explicitly propose to compute the *dynamic* SB with the additional time moments $t \in [0, 1]$ via the score-based diffusion models to learn the desired optimal path, under different specific restrictions (e.g., continuous v.s. discrete data space and unpaired v.s. paired domain translation). These prior works largely target static domains in practical applications, leaving open the question of how DSB performs in settings where the dynamic trajectory is physically meaningful, which is a question we directly investigate in this work.

2.2 Inverse astrostatistical predictions for star formation

The real-world physical problem we address in this paper focuses on a key class of inverse prediction tasks in astrostatistics, particularly within the context of star formation (McKee and Ostriker, 2007; Larson, 2003; Madau and Dickinson, 2014). We focus on the *dynamic physical processes* within Giant Molecular Clouds (GMCs) (Shu et al., 1987), a major component of the interstellar medium

³Terminology note: The term “*paired DSB*” is used here in the ML sense, which refers to the sample-level conditioning between observables and physical states, rather than in the classical unpaired Schrödinger Bridge formulation based on probabilistic coupling.

(ISM) (Spitzer Jr, 2008) that directly regulates star formation rates. The goal of our task is to infer true physical distributions, such as density or magnetic fields, from partial observables. In contrast to image generation or translation tasks where input-output mappings are static and visually structured, our problem setup reflects dynamic and physically driven distributional transitions.

Traditionally in astronomy, such observational inverse prediction problems have been commonly approached via *astrostatistical methods*, which often combine theoretical simplifications with statistical priors (also in Sec. 4). For instance, molecular density is commonly estimated using two- or three-component power-law conversion derived under idealized physical assumptions about emission properties and cloud structure (Bisbas et al., 2021, 2023). While effective within limited regimes, these methods tend to lack generalization ability when applied to systems governed by varying physical conditions and dynamics, making their performance unreliable in unseen testing cases.

Grounding DSB modeling within this context thus enables a more meaningful evaluation of probabilistic models beyond synthetic benchmarks (Liu et al., 2023; Gushchin et al., 2024; Ksenofontov and Korotin, 2025; Su et al., 2023). Notably, we leverage a key strength of the astrophysical setting: the ability to define *Out-of-Distribution (OOD)* conditions in a principled and physically grounded way through simulations with varying initial conditions and dominant physical processes. In addition to standard evaluation on *In-Domain (ID)* cases, our experiments explicitly assess OOD generalization, offering new insights into the robustness of probabilistic modeling in scientific prediction tasks.

3 Diffusion Schrödinger bridge for astrophysical observational inversion

3.1 Problem statement and preliminary in astronomy and ML

GMCs observational prediction. Studying the precise physical distributions within Giant Molecular Clouds (GMCs) is challenging yet essential for understanding the star formation process in astronomy. This challenge mainly arises from the highly complex and nonlinear physical dynamics within the forward evolution and the observational constraints.

Given a physical initial condition $\mathbf{x}_0$, we assume that the GMCs evolve under a set of governing physical laws as modeled in previous astrophysical studies (Wu et al., 2015, 2020; Hsu et al., 2023). This results in an equilibrium physical state $\mathbf{x}_1$, which we denote as $\mathbf{x}_1 = \mathcal{H}(\mathbf{x}_0)$, where $\mathcal{H}$ is a highly complex, nonlinear function incorporating self-gravity, magnetic fields, radiative heating, and cooling processes. While the forward process $\mathcal{H}$ is typically well-defined in controlled simulations, direct access to $\mathbf{x}_1$ is *rarely available in practice*. Instead, we often only observe a limited projection $\mathbf{y}$ of the true physical state. Therefore, our objective is to infer the full distribution $\mathbf{x}_1$ from partial observable $\mathbf{y}$, featuring a class of observational inverse prediction problems that lie at the core of many tasks in astrostatistical modeling. We further assume an observational model in the form of:

$$\mathbf{y} = \mathcal{F}(\mathbf{x}_1) + \varepsilon = \mathcal{F}(\mathcal{H}(\mathbf{x}_0)) + \varepsilon, \quad (1)$$

where $\mathcal{F}$ represents the observational process (e.g., line-of-sight integration, instrumental response) and ε denotes corresponding observational noise in the measurement. Given this formulation, our core task becomes a conditional probabilistic inference problem, where the goal is to estimate the posterior distribution $p(\mathbf{x}_1|\mathbf{y})$.

Preliminary on dynamic DSB. Consider two distributions p_0, p_1 and a reference path measure $\mathbb{Q} \in \mathcal{P}(\mathcal{C})$, the dynamic SB problem seeks to find a path measure $\mathbb{P}^{SB} \in \mathcal{P}(\mathcal{C})$ such that:

$$\mathbb{P}^{SB} = \operatorname{argmin}_{\mathbb{P}} \{\text{KL}(\mathbb{P}|\mathbb{Q}) : \mathbb{P}_0 = p_0, \mathbb{P}_1 = p_1\}, \quad (2)$$

with $\mathcal{C} = C([0, 1], \mathbb{R}^d)$ defined as the space of continuous functions from time space $[0, 1]$ to $\mathbb{R}^d$. SB can be formalized as an entropy-regularized OT problem describing the following forward and backward SDEs:

$$\begin{aligned} d\mathbf{x}_t &= [f(\mathbf{x}, t) + g(t)^2 \nabla \log \Psi(\mathbf{x}, t)]dt + g(t)d\mathbf{w}, \\ d\mathbf{x}_t &= [f(\mathbf{x}, t) - g(t)^2 \nabla \log \hat{\Psi}(\mathbf{x}, t)]dt + g(t)d\bar{\mathbf{w}}. \end{aligned} \quad (3)$$

Here, $f(\cdot, t)$ is the linear drift, $g(t)$ controls the diffusion intensity over time, and $d\mathbf{w}$ represents the standard Wiener processes for modeling stochastic noise. The wave functions Ψ and $\hat{\Psi}$ are

time-varying energy potentials that solve the following coupled PDEs:

$$\begin{aligned}\frac{\partial \Psi(\mathbf{x}, t)}{\partial t} &= -\nabla \Psi(\mathbf{x}, t)^T f(\mathbf{x}, t) - \frac{1}{2}g(t)^2 \Delta \Psi(\mathbf{x}, t), \\ \frac{\partial \hat{\Psi}(\mathbf{x}, t)}{\partial t} &= \nabla \cdot (\hat{\Psi}(\mathbf{x}, t) f(\mathbf{x}, t) + \frac{1}{2}g(t)^2 \Delta \hat{\Psi}(\mathbf{x}, t),\end{aligned}\tag{4}$$

with $\Psi(\mathbf{x}, 0)\hat{\Psi}(\mathbf{x}, 0) = p_0(\mathbf{x})$ and $\Psi(\mathbf{x}, 1)\hat{\Psi}(\mathbf{x}, 1) = p_1(\mathbf{x})$. Recent works have shown that score-based diffusion models provide a practical numerical solver for the DSB problem by parameterizing the time-dependent scores associated with forward and backward SDEs (De Bortoli et al., 2021; Shi et al., 2023; Liu et al., 2023). Specifically, instead of explicitly solving the Schrödinger system via wave potentials Ψ and $\hat{\Psi}$, SGMs learn the forward and reverse score functions $\nabla \log p_t(\mathbf{x})$ through denoising score matching, simulating the optimal stochastic process by integrating the learned SDEs. This formulation enables the DSB problem to be learned from data via stochastic sampling and bypasses the need for explicit access to density functions $p_t(\mathbf{x})$, which are often intractable in real-world scientific applications. In our work, we build on this framework and introduce a pairwise-constrained variant of the DSB model tailored to astrophysical observational inversion tasks, as described in the following section.

3.2 Diffusion Schrödinger bridge for astrophysical observational prediction (Astro-DSB)

Core technical motivation and relation to prior works. Our core technical motivation seeks to better align the states $\mathbf{x}_t$ in score-based diffusion models with the dynamic evolution processes governing GMCs, which leads to two key differences from existing literature.

Within this field of observational prediction in GMCs, existing state-of-the-art (SOTA) ML-based methods (Xu et al., 2023a,b, 2025) leverage conditional diffusion denoising probabilistic models (DDPMs), framing the observational data $\mathbf{y}$ as the condition. While achieving superior performance compared to traditional astrostatistical methods, conditional DDPMs *introduce unnecessary Gaussian prior assumptions*, which are not naturally aligned with the physics of GMCs. Our proposed Schrödinger bridge framework removes this assumption by directly modeling the transition between observable $\mathbf{y}$ and the true physical distribution $\mathbf{x}_1$, resulting in a distribution-level probabilistic mapping *without reliance on artificial prior structures*, thus offering better ML interpretability compared to conditional DDPMs.

From the physics-aligned modeling perspective, Li et al. (2025) recently propose another variant of diffusion bridge with a two-stage learning process (data-driven pretraining and physics-aligned finetuning) for physical field reconstruction and super-resolution. In contrast, our approach adopts a single-stage learning framework to learn the transition path directly without pretraining, and systematically evaluates both in-distribution performance and *OOD generalization* to assess robustness, further extending the prediction to real observational data.

M0: Pairwise matching in marginal distributions. Similar to many other physical tasks (Li et al., 2025), our observational prediction problem exhibits a pairwise connection between marginal distributions p_0 and p_1 established by $\mathbf{y}$ and $\mathbf{x}_1$, respectively. Based on this, we follow the prior work I²SB (Liu et al., 2023) for a special case of the SB problem. Specifically, I²SB presents a tractable class of SB that eliminates the couplings of the wave functions from Eq. 3, and propose an analytical form for the implementation of the posterior distribution:

$$p(\mathbf{x}_t | \mathbf{y}, \mathbf{x}_1) = \mathcal{N}(\mathbf{x}_t; \frac{\bar{\sigma}_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} \mathbf{y} + \frac{\sigma_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} \mathbf{x}_1, \frac{\bar{\sigma}_t^2 \sigma_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} \mathbf{I}),\tag{5}$$

where $\sigma_t^2 := \int_0^t \beta_\tau d\tau$ and $\bar{\sigma}_t^2 := \int_t^1 \beta_\tau d\tau$ denote accumulated variances. This forms the base design of our proposed *Astro-DSB*. However, we observe that this baseline framework alone is insufficient for accurately predicting true physical distributions from constrained observational inputs.

M1: Noise perturbation alignment between ML and observables. Compared to conventional image translation (Liu et al., 2023) and other physical tasks such as field reconstruction (Li et al., 2025) that take a clean data input $\mathbf{y}$, we note that explicitly modeling the noise perturbation term ϵ in observables $\mathbf{y}$ in Eq. 5 offers a simple yet critical alignment to physical assumptions in observational

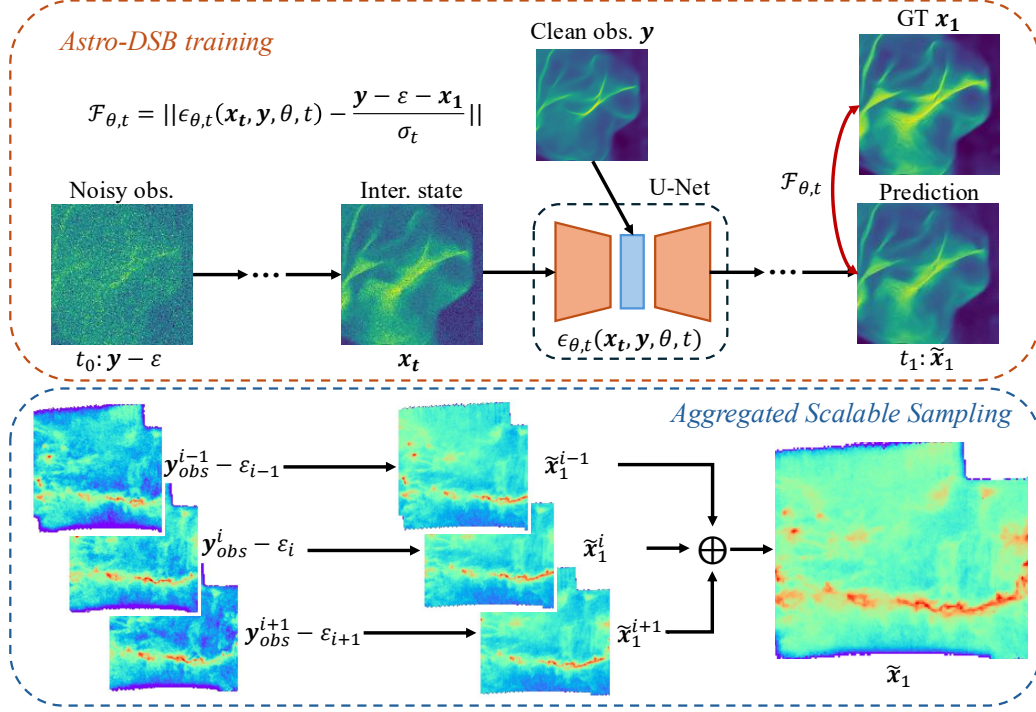


Figure 2: **Illustration pipeline of our proposed *Astro-DSB* method.** In the training stage, we propose to learn the DSB model under the pairwise matching assumption with observational noises alignment ϵ and enhancement observables $\mathbf{y}$. In inference, we adopt an aggregated scalable sampling strategy to infer physical distributions from larger observables, such as Taurus B213.

prediction scenarios described in Eq. 1. In this work, we assume that the noise introduced from the observation process follows a weighted Gaussian, denoted as $\epsilon \sim \lambda \mathcal{N}(0, I)$. Following statistical conventions in astrophysical studies, we set $\lambda = 0.1$ throughout the experiments. This modification ensures that our *Astro-DSB* framework respects the uncertainty structure of constrained observables, refining the targeted conditional posterior distribution to be $p(\mathbf{x}_1 | \mathbf{y} - \epsilon)$.

M2: Observable enhancement within DSB model. Compared to the vanilla I^2SB framework, we further propose to enhance the learning process by incorporating the clean observable input $\mathbf{y}$ as an additional intermediate condition at each time step t . This modification aims to facilitate the learning of the score function $\epsilon_{\theta,t}$, implemented through U-Net architecture (Ronneberger et al., 2015), by consistently “reminding” the model of the observational constraint throughout the trajectory. Empirically, we find this design critical for achieving robust prediction performance within the DSB framework and helps simultaneously obtain lower loss values with a faster convergence speed.

Patch-based training and aggregated scalable sampling. Following the methodological designs described above, we train the proposed *Astro-DSB* using the reparameterized objective function:

$$\mathcal{F}_{\theta,t} = \left\| \epsilon_{\theta,t}(\mathbf{x}_t, \mathbf{y}, t; \theta) - \frac{\mathbf{y} - \epsilon - \mathbf{x}_1}{\sigma_t} \right\|^2, \quad (6)$$

where $\mathbf{x}_t$ takes the analytical form from Eq. 5. Given that astrophysical data often present large spatial resolutions, we adopt a patch-based training strategy that is widely used in both vision and scientific applications (Li et al., 2025; Yang et al., 2019), with cropped patches from the original simulation data. Specifically, $\mathbf{x}_1$ and $\mathbf{y}$ are cropped into smaller patches of consistent size to facilitate efficient learning while preserving local physical structures.

For inference, in contrast to the padding-based techniques used in prior works (Li et al., 2025), we design an aggregated sampling strategy to restore higher-resolution prediction results. Given a large testing observable $\mathbf{y}_{obs}$ (e.g., real-world survey data), we partition $\mathbf{y}_{obs}$ into overlapping patches

$\mathbf{y}_{obs}^i$ that match the training resolution. We then independently run inference on each patch to obtain corresponding predictions $\tilde{\mathbf{x}}_1^i$, and aggregate across all patches to reconstruct the final predicted distribution $\tilde{\mathbf{x}}_1$. Fig. 2 illustrates the described pipeline, with detailed algorithms in the appendices.

4 Experiments

4.1 Astrophysical data

Magnetohydrodynamics simulations. Similar to previous work (Xu et al., 2023b, 2025), our training data are drawn from synthetic magnetohydrodynamics (MHD) simulations. Specifically, we follow the simulation setup from Wu et al. (2020) and Hsu et al. (2023), which are conducted using the MUSCL-Dedner method and HLLD Riemann solver in the adaptive mesh refinement (AMR) code Enzo (Dedner et al., 2002; Wang and Abel, 2009; Bryan et al., 2014). These simulations model key physical processes including self-gravity, magnetic fields, radiative heating, and cooling, based on a photodissociation model (Wu et al., 2015). We consider the *molecular density* (ρ , measured in the number of H nuclei n_H/cm^{-3}) and magnetic field strength (B , measured in μG) as the primary quantities for prediction in this work. For molecular density, we collected data from 7179 simulation trials, split into 5707 training and 1472 testing cases following an 8:2 ratio. For magnetic field strength, we obtained 19100 training and 6370 testing cases. All data are pre-processed by a logarithmic transformation followed by normalization to $[0,1]$, and resized to 128×128 patches.

Additionally, we generate 1680 independent tests with different dominate physical process assumptions as the OOD testing cases to evaluate generalization. Specifically, unlike the ID MHD simulations code ENZO (Bryan et al., 2014) initialized with colliding and non-colliding molecular clouds, OOD data are produced using a different MHD code, ORION2 (Li et al., 2021), which excludes gravity but includes driven turbulence, and employs a different set of initial conditions. While the fundamental MHD equations are similar between two simulation codes, the physical processes and initial setups differ significantly, thus clearly distinguishing the ID and OOD cases under physically meaning shifts.

Synthetic and real observables. After generating the simulation data, we define the observables $\mathbf{y}$ following conventions in astrophysical studies. For molecular density, we adopt the projected mass surface density maps (also known as the “column density”, measured in N_H/cm^{-2}) as the observables $\mathbf{y}_\rho$. For the magnetic field, the synthetic observable $\mathbf{y}_B$ consists of the column density, the orientation angles of the dust continuum polarization vector, and line-of-sight (LOS) nonthermal velocity dispersion. In addition to these synthetic observables, we also evaluate the proposed *Astro-DSB* on real observational data from the Taurus B213 region, using column density maps provided by Palmeirim et al. (2013). B213 is one of closest star-forming filamentary structures located in the Taurus molecular cloud (Li and Goldsmith, 2012). Unlike the synthetic data where ground-truth distributions are available, the Taurus B213 observation $\mathbf{y}_{B213}$ lacks a precise reference for physical states, and is instead constrained by indirect measurements from complementary observations. This real-data evaluation demonstrates the applicability of *Astro-DSB* in practical scenarios where ground truth is unavailable, and highlights its potential for scientific discovery in observational settings.

4.2 Comparing methods and evaluations

Astrostatistical methods. The conventional methods involve astrostatistical techniques to infer physical distributions with simplified physical assumptions. For molecular density prediction, existing popular techniques include statistical conversions such as 2-component (2PLF) and 3-component power-law fitting (3PLF). For magnetic field, Davis–Chandrasekhar–Fermi (DCF) method (Davis, 1951; Chandrasekhar and Fermi, 1953; Beck, 2016) has been the classic method for estimating the magnetic field strength, which relies on the assumption of an equipartition between the magnetic field energy and the turbulent kinetic energy of the gas.

Machine learning methods. Recent studies have also adopted machine learning techniques, which can be broadly categorized into discriminative and generative approaches. Discriminative models such as convolutional neural networks (CNNs) trained with reconstruction losses have been widely adopted (Xu et al., 2020a,b). From the generative modeling perspective, conditional Denoising Diffusion Probabilistic Models (DDPMs) (Ho et al., 2020) have emerged as a recent popular alternative to tackle those observational prediction problems.

Table 1: **Quantitative evaluation of density prediction results reported on the original data scale ($10^{0\sim7}$).** The signs indicate the direction of prediction bias: negative denotes overprediction, and positive indicates underprediction. We **bold** the best results and underline the second-best. μ and σ are the primary criteria for assessment, where values closer to zero indicate superior performance.

ρ (n_H/cm^{-3})		ID		OOD		
Methods	$\mu - p_{\Delta\rho/\rho}$ ($\downarrow\downarrow$)	$\sigma - p_{\Delta\rho/\rho}$ ($\downarrow\downarrow$)	ME	$\mu - p_{\Delta\rho/\rho}$ ($\downarrow\downarrow$)	$\sigma - p_{\Delta\rho/\rho}$ ($\downarrow\downarrow$)	ME
2PLF	-2.50 $\pm$ 7E-4	7.11 $\pm$ 2E-3	-715.95 $\pm$ 1.88	-5.77 $\pm$ 3E-4	7.21 $\pm$ 1E-3	-1680.82 $\pm$ 0.74
3PLF	-0.77 $\pm$ 5E-4	4.44 $\pm$ 2E-3	-715.78 $\pm$ 1.88	-3.23 $\pm$ 2E-4	4.21 $\pm$ 7E-4	-1680.60 $\pm$ 0.61
U-NET	-0.25 $\pm$ 1E-4	1.29 $\pm$ 4E-4	118.24 $\pm$ 0.94	5.01 $\pm$ 6E-4	7.08 $\pm$ 1E-3	1313.60 $\pm$ 0.61
cDDPMs	-0.05 $\pm$ 6E-5	0.61$\pm$6E-4	4.61 $\pm$ 1.05	<u>2.88$\pm$5E-4</u>	<u>5.33$\pm$7E-4</u>	1050.51 $\pm$ 0.61
Astro-DSB (Ours)	-0.02$\pm$1E-4	<u>0.71$\pm$3E-4</u>	32.09 $\pm$ 0.65	0.51$\pm$2E-4	2.32$\pm$3E-3	-5.04 $\pm$ 1.32

In our experiments, we compare with both conventional and ML-based methods. Specifically for ML, in addition to the state-of-the-art conditional DDPMs, we also re-implement a U-Net (Ronneberger et al., 2015) from CASI-2D (Xu et al., 2020a) as a pixel-to-pixel approach for comparison. Notably, while recent works suggest that generative models offer improved sample quality in image synthesis, their advantage in scientific prediction, especially under physically grounded distribution shift (OOD), remains unclear. Our study explicitly tests this hypothesis by comparing probabilistic and pixel-to-pixel modeling under controlled simulation setups with similar model scales.

Evaluations. Our evaluation adheres to established protocols in astrophysical studies, emphasizing **the distributional patterns of relative error between predictions and ground truth values**. Specifically, we assess the normalized relative error (e.g., $\Delta\rho/\rho$), where an ideal prediction should produce a distribution centered around zero. Although this distribution may not strictly follow a Gaussian form, a symmetric shape with a narrow spread around zero indicates accurate and well-calibrated predictions. In contrast, distributions with large bias or heavy tails reflect systematic under- or over-estimation. To quantitatively capture these characteristics, we report the mean μ (gt - pred) and standard deviation σ of the relative error distribution as **primary metrics**. These values provide insights into the central tendency and variability of model performance across different test cases.

In addition to relative error metrics, we also report the Mean Error (ME) between predicted and ground truth values. However, it is important to note that ME interpretation may be affected by the wide dynamic range of physical quantities in astrophysical studies. Specifically, **even a method with a lower ME value does not necessarily indicate superior performance**, as the absolute error can be dominated by regions of large values. For example, a small relative error on a value of 10^7 can contribute disproportionately to the ME compared to a larger relative error on a value of 10. This imbalance highlights the importance of relative error patterns in accurately assessing performance.

Implementation details. To ensure a fair comparison among the tested ML methods, we fix the U-Net backbone architecture across the discriminative U-Net, conditional DDPMs, and our proposed *Astro-DSB*, resulting in approximately 81M learnable parameters for each model. For the U-Net baseline, we train the model for 20 epochs using Mean Square Error (MSE) loss. For conditional DDPMs, we follow the setup from prior works (Xu et al., 2023a, 2025) and adopt a 1000-step schedule with the standard variational lower bound loss (Ho et al., 2020). Our proposed *Astro-DSB* is trained with the same time discretization (1000 steps) using the objective defined in Eq. 5. All models are trained with a batch size of 16 using the AdamW optimizer with a learning rate of 5e-5, running on an AWS cluster with 4 NVIDIA T4 GPUs. Notably, while conditional DDPMs require around 400 epochs to converge, *Astro-DSB* converges in only around 100 epochs under the same setup, reducing overall training time by approximately 75%. This aligns with our expectations and observations in prior vision DSB works, as the target distribution p_1 now lies closer to the physically meaningful observation distribution p_0 , compared to the case where p_0 is a zero-mean Gaussian in DDPMs. During inference, the runtime per testing case ranges from approximately 6 to 30 seconds, depending on whether the step skipping (Song et al., 2021) is applied. Notably, we observe no significant difference in prediction performance across these sampler settings in our work.

4.3 Experimental results and analysis

Less biased prediction results for ID and OOD testing. We present the quantitative evaluation results for density and magnetic field prediction in Tab. 1 and Tab. 2, respectively. Our proposed *Astro-DSB* consistently achieves the best or second-best results across both ID and OOD settings,

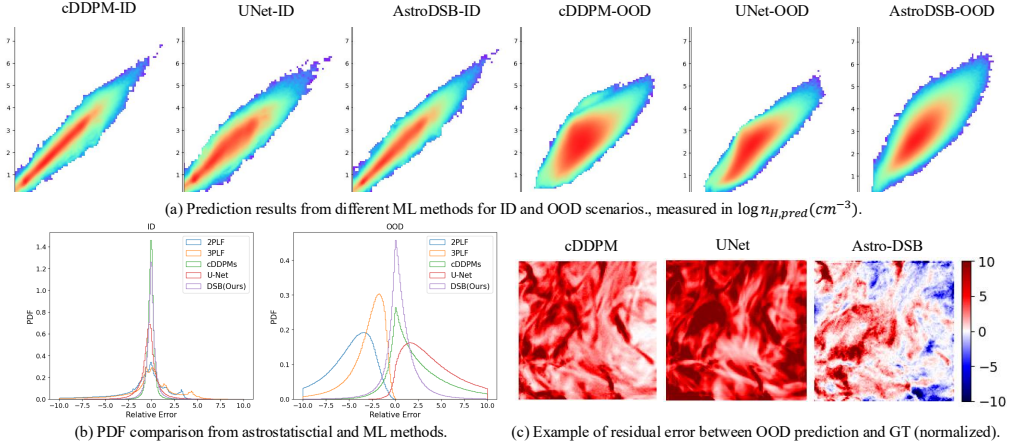


Figure 3: **Multiple qualitative visualizations for relative error comparisons.** (a) 2D histogram showing the distribution of ground truth (x-axis) versus predictions (y-axis). An accurate and unbiased prediction lies along the 1-to-1 diagonal line. (b) Relative error between the ground truth and predictions. (c) Residual maps showing the difference between model predictions and the ground truth across different methods. Results from *Astro-DSB* present more balanced and unbiased prediction error patterns compared to other ML methods, particularly for OOD testing cases. Best viewed in color and with zoom-in.

Table 2: **Quantitative evaluations of magnetic field prediction results reported on the original data scale ($3\text{E}^{1\sim3}$).** **Bold** values indicate the best results, underlined values indicate the second-best. *Astro-DSB* achieves competitive SOTA with only 25% of the computational cost of cDDPMs.

B (μG)		ID			OOD		
Methods	$\mu - p_{\Delta B/B}$ ($\downarrow$)	$\sigma - p_{\Delta B/B}$ ($\downarrow$)	ME	$\mu - p_{\Delta B/B}$ ($\downarrow$)	$\sigma - p_{\Delta B/B}$ ($\downarrow$)	ME	
DCF	$11.80 \pm 8\text{E-}4$	$21.78 \pm 4\text{E-}4$	507.74 ± 0.03	$5.19 \pm 8\text{E-}4$	$9.64 \pm 3\text{E-}3$	137.76 ± 0.03	
PLF	$-0.11 \pm 5\text{E-}5$	$1.23 \pm 1\text{E-}4$	$-1.04 \pm 1\text{E-}3$	$1.03 \pm 9\text{E-}5$	$1.47 \pm 1\text{E-}4$	$17.14 \pm 2\text{E-}3$	
U-NET	$-0.06 \pm 2\text{E-}5$	$0.38 \pm 5\text{E-}5$	154.76 ± 0.08	$-0.32 \pm 6\text{E-}5$	$0.74 \pm 7\text{E-}5$	146.38 ± 0.02	
cDDPMs	<u>$0.07 \pm 2\text{E-}5$</u>	$0.38 \pm 6\text{E-}5$	123.99 ± 0.04	$0.12 \pm 7\text{E-}5$	<u>$0.72 \pm 2\text{E-}4$</u>	150.00 ± 0.04	
Astro-DSB (Ours)	<u>$0.15 \pm 2\text{E-}5$</u>	<u>$0.55 \pm 5\text{E-}5$</u>	235.85 ± 0.08	<u>$0.19 \pm 6\text{E-}5$</u>	$0.67 \pm 9\text{E-}5$	131.89 ± 0.02	

particularly demonstrating robust superiority in OOD predictions, as indicated by the lower mean (μ) and standard deviation (σ) of the relative error distribution. In contrast, traditional astrostatistical methods such as power-law fitting (PLF) and DCF exhibit large error values, while U-Net and DDPMs suffer from more severe performance degradation under OOD conditions. In particular, the training cost of *Astro-DSB* is only 25% of the previous SOTA conditional DDPMs method.

Fig. 3 further highlights the superior generalization of *Astro-DSB*, with sharply peaked PDF curves and coherent residual error patterns, especially in OOD scenarios. In comparison, both machine learning methods, U-Net modeling and cDDPMs, tend to underpredict in OOD testing, while conventional methods, including 2PL and 3PL, usually overpredict for density estimations in unseen simulations. These results demonstrate that *Astro-DSB* offers robust and efficient predictions for astrophysical inverse tasks. Extended evaluations can be found in the appendices.

Extended robustness on observations from Taurus B213. We further test our *Astro-DSB* on the Herschel column density map of Taurus B213 (Li and Goldsmith, 2012; Palmeirim et al., 2013) using the model trained purely on synthetic simulations. Fig. 4 visualizes the observed column density map alongside our predicted volume density map. Despite the lack of pixel-level ground truth in real-world observations, our predictions demonstrate a coherent spatial structure and effectively capture the filamentary features of B213, as qualitatively verified by astrophysicists. Notably, the results align well with single-pointing constraints derived from indirect measurements using the density tracer HC_3N (Li and Goldsmith, 2012), which suggests that *Astro-DSB* can generalize effectively beyond simulated scenarios, providing reliable physical predictions even on real observational data. More experimental results are in the appendices.

Table 3: Ablation results for the key methodology designs on density predictions.

$\rho (n_H)$		ID		OOD		
Method	$\mu - p_{\Delta\rho/\rho} (\downarrow)$	$\sigma - p_{\Delta\rho/\rho} (\downarrow)$	ME	$\mu - p_{\Delta\rho/\rho} (\downarrow)$	$\sigma - p_{\Delta\rho/\rho} (\downarrow)$	ME
w/o M1 & M2	$-0.19 \pm 3E-4$	1.75 ± 0.04	66.54 ± 1.49	$4.24 \pm 2E-3$	14.71 ± 0.12	1086.66 ± 0.66
w/o M1	$-3E-3 \pm 7E-5$	$0.65 \pm 6E-4$	35.86 ± 0.44	$-1.84 \pm 5E-4$	$5.06 \pm 4E-3$	-2297.48 ± 1.10
w/o M2	$7E-3 \pm 1E-4$	$1.63 \pm 5E-4$	37.15 ± 1.53	$-14.66 \pm 5E-3$	54.95 ± 3.47	-12069.44 ± 9.44
Astro-DSB (Full)	$-0.02 \pm 1E-4$	$0.71 \pm 3E-4$	32.09 ± 0.65	$0.51 \pm 2E-4$	$2.32 \pm 3E-3$	-5.04 ± 1.32

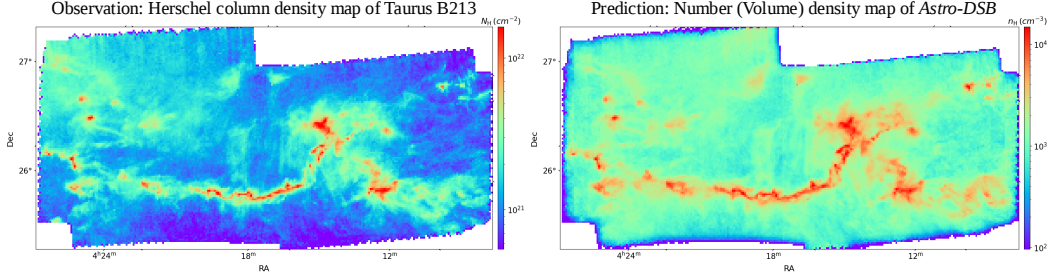


Figure 4: **Visualization of observables and our predicted results on Herschel column density map of Taurus B213**, shown in the plane of the sky with the x-axis representing right ascension and the y-axis representing declination. The observable map is measured in N_H/cm^{-2} , and the prediction is in n_H/cm^{-3} . While there is no ground truth for real-world observations, our prediction results are consistent with those obtained with other independent approaches.

Ablation studies on key methodological components. We evaluate the effectiveness of two key components in our *Astro-DSB* design: **M1: Noise alignment** and **M2: Observable enhancement**, by systematically removing them and reporting the results in Tab. 3. Our findings reveal that the absence of either module significantly degrades performance, with **M2** being particularly effective for cases of OOD and stable prediction indicated as indicated by σ values. The full *Astro-DSB* achieves the best balance of prediction accuracy and generalization robustness.

5 Conclusion and Discussion

In this work, we address a class of astrostatistical inverse prediction tasks by introducing *Astro-DSB*, a dynamic generative modeling framework based on the Diffusion Schrödinger Bridge. Targeting key physical quantities within GMCs, *Astro-DSB* learns the probabilistic transition between partial observables and true physical states, offering a more flexible alternative to conventional one-step or purely discriminative approaches. Our study reveals that, in contrast to static tasks where dynamic modeling may appear excessive, explicitly learning distributional transitions is both meaningful and beneficial in physically grounded systems demonstrated via improved generalization in out-of-domain simulations and real-world observables. Overall, *Astro-DSB* offers better alignment with physical interpretability and supports the broader potential of probabilistic modeling in scientific discovery.

Limitation and future direction. Despite the improved robustness and performance to unseen simulation conditions, our method does not explicitly model the full forward physical equations that govern the GMC evolution. While this is reasonable given the intractable nature of observational inversion problems in this work, future extensions could involve integrating multiple intermediate states, leveraging richer supervision signals, or directly encoding the dynamic physical laws into the generative modeling process to further strengthen the alignment between ML and physical dynamics for extended applications with well-defined physical forward processes.

Acknowledgments and Disclosure of Funding

This research was partially supported by an Amazon Research Award to OR and YZ. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of Amazon. DX acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number 568580]. DX also acknowledges support from the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship Program, a program of Schmidt Sciences. YZ also acknowledges the travel funding support by the French National Research Agency (ANR) via the “GraspGNNs” JCJC grant (ANR-24-CE23-3888), coordinated by Johannes F. Lutzeyer from École Polytechnique.

References

- Rainer Beck. Magnetic fields in spiral galaxies. *The Astronomy and Astrophysics Review*, 24(1):4, 2016.
- Thomas G Bisbas, Jonathan C Tan, and Kei EI Tanaka. Photodissociation region diagnostics across galactic environments. *Monthly Notices of the Royal Astronomical Society*, 502(2):2701–2732, 2021.
- Thomas G Bisbas, Ewine F van Dishoeck, Chia-Yu Hu, and Andreas Schruba. Pdfchem: A new fast method to determine ism properties and infer environmental parameters using probability distributions. *Monthly Notices of the Royal Astronomical Society*, 519(1):729–753, 2023.
- Greg L. Bryan, Michael L. Norman, Brian W. O’Shea, Tom Abel, John H. Wise, Matthew J. Turk, Daniel R. Reynolds, David C. Collins, Peng Wang, Samuel W. Skillman, Britton Smith, Robert P. Harkness, James Bordner, Ji-hoon Kim, Michael Kuhlen, Hao Xu, Nathan Goldbaum, Cameron Hummels, Alexei G. Kritsuk, Elizabeth Tasker, Stephen Skory, Christine M. Simpson, Oliver Hahn, Jeffrey S. Oishi, Geoffrey C. So, Fen Zhao, Renyue Cen, Yuan Li, and Enzo Collaboration. ENZO: An Adaptive Mesh Refinement Code for Astrophysics. *The Astrophysical Journal Supplement Series*, 211(2):19, 2014.
- Greg L Bryan, Michael L Norman, Brian W O’Shea, Tom Abel, John H Wise, Matthew J Turk, Daniel R Reynolds, David C Collins, Peng Wang, Samuel W Skillman, et al. Enzo: An adaptive mesh refinement code for astrophysics. *The Astrophysical Journal Supplement Series*, 211(2):19, 2014.
- Subrahmanyan Chandrasekhar and Enrico Fermi. Magnetic fields in spiral arms. *Astrophysical Journal*, 118: 113–115, 1953.
- Richard M Crutcher. Magnetic fields in molecular clouds: observations confront theory. *The Astrophysical Journal*, 520(2):706, 1999.
- Richard M Crutcher. Magnetic fields in molecular clouds. *Annual Review of Astronomy and Astrophysics*, 50(1): 29–63, 2012.
- Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.
- Leverett Davis. The strength of interstellar magnetic fields. *Phys. Rev.*, 81:890–891, 1951.
- Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schrödinger bridge with applications to score-based generative modeling. *NeurIPS*, 34:17695–17709, 2021.
- Andreas Dedner, Friedemann Kemm, Dietmar Kröner, C-D Munz, Thomas Schnitzer, and Matthias Wesenberg. Hyperbolic divergence cleaning for the mhd equations. *Journal of Computational Physics*, 175(2):645–673, 2002.
- Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. In *NeurIPS*, 2021.
- Kieran Didi, Francisco Vargas, Simon V Mathis, Vincent Dutordoir, Emile Mathieu, Urszula J Komorowska, and Pietro Lio. A framework for conditional diffusion modelling with applications in motif scaffolding for protein design. *arXiv preprint arXiv:2312.09236*, 2023.
- Nikita Gushchin, Daniil Selikhanovych, Sergei Kholkin, Evgeny Burnaev, and Alexander Korotin. Adversarial schrödinger bridge matching. *NeurIPS*, 2024.
- Mark Heyer and TM Dame. Molecular clouds in the milky way. *Annual Review of Astronomy and Astrophysics*, 53(1):583–629, 2015.
- Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, 2020.

- Jonathan Ho, Chitwan Saharia, William Chan, David J Fleet, Mohammad Norouzi, and Tim Salimans. Cascaded diffusion models for high fidelity image generation. *Journal of Machine Learning Research*, 2022.
- Chia-Jung Hsu, Jonathan C Tan, Duncan Christie, Yu Cheng, and Theo J O’Neill. Gmc collisions as triggers of star formation–viii. the core mass function. *Monthly Notices of the Royal Astronomical Society*, 522(1): 700–720, 2023.
- Mark R Krumholz, Christopher F McKee, and Joss Bland-Hawthorn. Star clusters across cosmic time. *Annual Review of Astronomy and Astrophysics*, 57(1):227–303, 2019.
- Grigoriy Ksenofontov and Alexander Korotin. Categorical schrödinger bridge matching. *arXiv preprint arXiv:2502.01416*, 2025.
- Richard B Larson. The physics of star formation. *Reports on Progress in Physics*, 66(10):1651, 2003.
- Di Li and Paul F Goldsmith. Is the taurus b213 region a true filament?: Observations of multiple cyanoacetylene transitions. *The Astrophysical Journal*, 756(1):12, 2012.
- Pak Li, Andrew Cunningham, Brandt Gaches, Richard Klein, Mark Krumholz, Aaron Lee, Christopher McKee, Stella Offner, Anna Rosen, and Aaron Skinner. ORION2: A magnetohydrodynamics code for star formation. *The Journal of Open Source Software*, 6(68):3771, 2021.
- Zeyu Li, Hongkun Dou, Shen Fang, Wang Han, Yue Deng, and Lijun Yang. Physics-aligned field reconstruction with diffusion bridge. In *ICLR*, 2025.
- Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos A Theodorou, Weili Nie, and Anima Anandkumar. I²sb: Image-to-image schrödinger bridge. *ICML*, 2023.
- Piero Madau and Mark Dickinson. Cosmic star-formation history. *Annual Review of Astronomy and Astrophysics*, 52(1):415–486, 2014.
- Thomas Maschberger and Pavel Kroupa. Estimators for the exponent and upper limit, and goodness-of-fit tests for (truncated) power-law distributions. *Monthly Notices of the Royal Astronomical Society*, 395(2):931–942, 2009.
- Christopher F McKee and Eve C Ostriker. Theory of star formation. *Annu. Rev. Astron. Astrophys.*, 45(1): 565–687, 2007.
- P al Palmeirim, Ph André, J Kirk, Derek Ward-Thompson, Det al Arzoumanian, V Könyves, P Didelon, N Schneider, M Benedettini, Sylvain Bontemps, et al. Herschel view of the taurus b211/3 filament and striations: evidence of filamentary growth? *Astronomy & Astrophysics*, 550:A38, 2013.
- Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, 2022.
- Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015.
- Erwin Schrödinger. Sur la théorie relativiste de l’électron et l’interprétation de la mécanique quantique. In *Annales de l’institut Henri Poincaré*, pages 269–310, 1932.
- Yuyang Shi, Valentin De Bortoli, George Deligiannidis, and Arnaud Doucet. Conditional simulation using diffusion schrödinger bridges. In *Uncertainty in Artificial Intelligence*, pages 1792–1802. PMLR, 2022.
- Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion schrödinger bridge matching. *NeurIPS*, 36:62183–62223, 2023.
- Frank H Shu, Fred C Adams, and Susana Lizano. Star formation in molecular clouds-observation and theory. *IN: Annual review of astronomy and astrophysics. Volume 25 (A88-13240 03-90). Palo Alto, CA, Annual Reviews, Inc., 1987, p. 23-81. NASA-sponsored research.*, 25:23–81, 1987.
- Raphael Skalidis and Konstantinos Tassis. High-accuracy estimation of magnetic field strength in the interstellar medium from dust polarization. *Astronomy & Astrophysics*, 647:A186, 2021.
- Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *ICML*. PMLR, 2015.

- Vignesh Ram Somnath, Matteo Pariset, Ya-Ping Hsieh, Maria Rodriguez Martinez, Andreas Krause, and Charlotte Bunne. Aligned diffusion schrödinger bridges. In *Uncertainty in Artificial Intelligence*, pages 1985–1995. PMLR, 2023.
- Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021.
- Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *ICLR*, 2020.
- Lyman Spitzer Jr. *Physical processes in the interstellar medium*. John Wiley & Sons, 2008.
- Xuan Su, Jiaming Song, Chenlin Meng, and Stefano Ermon. Dual diffusion implicit bridges for image-to-image translation. *ICLR*, 2023.
- Gefei Wang, Yuling Jiao, Qian Xu, Yang Wang, and Can Yang. Deep generative learning via schrödinger bridge. In *ICML*, pages 10794–10804. PMLR, 2021.
- Peng Wang and Tom Abel. Magnetohydrodynamic simulations of disk galaxy formation: the magnetization of the cold and warm medium. *The Astrophysical Journal*, 696(1):96, 2009.
- Benjamin Wu, Sven Van Loo, Jonathan C Tan, and Simon Bruderer. Gmc collisions as triggers of star formation. i. parameter space exploration with 2d simulations. *The Astrophysical Journal*, 811(1):56, 2015.
- Benjamin Wu, Jonathan C Tan, Duncan Christie, and Fumitaka Nakamura. Gmc collisions as triggers of star formation. vii. the effect of magnetic field strength on star formation. *The Astrophysical Journal*, 891(2):168, 2020.
- Duo Xu, Stella SR Offner, Robert Gutermuth, and Colin Van Oort. Application of convolutional neural networks to identify protostellar outflows in co emission. *The Astrophysical Journal*, 905(2):172, 2020a.
- Duo Xu, Stella SR Offner, Robert Gutermuth, and Colin Van Oort. Application of convolutional neural networks to identify stellar feedback bubbles in co emission. *The Astrophysical Journal*, 890(1):64, 2020b.
- Duo Xu, Stella SR Offner, Robert Gutermuth, Michael Y Grudić, Dávid Guszejnov, and Philip F Hopkins. Predicting the radiation field of molecular clouds using denoising diffusion probabilistic models. *The Astrophysical Journal*, 958(1):97, 2023a.
- Duo Xu, Jonathan C Tan, Chia-Jung Hsu, and Ye Zhu. Denoising diffusion probabilistic models to predict the density of molecular clouds. *The Astrophysical Journal*, 950(2):146, 2023b.
- Duo Xu, Jenna Karcheski, Chi-Yan Law, Ye Zhu, Chia-Jung Hsu, and Jonathan C Tan. Exploring magnetic fields in molecular clouds through denoising diffusion probabilistic models. *The Astrophysical Journal*, 980(1):52, 2025.
- Wenming Yang, Xuechen Zhang, Yapeng Tian, Wei Wang, Jing-Hao Xue, and Qingmin Liao. Deep learning for single image super-resolution: A brief review. *IEEE Transactions on Multimedia*, 21(12):3106–3121, 2019.
- Yujiao Yang, Chao Liu, Ming Yang, Yun Zheng, and Hao Tian. The star formation history of nearby galaxies: A machine learning approach. *The Astrophysical Journal*, 977(1):11, 2024.

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [\[Yes\]](#)

Justification: Our claims in the abstract and introduction are accurate reflections of the paper's contributions and scope.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [\[Yes\]](#)

Justification: We explicitly discuss the limitations and potential future directions in Section 5.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: We do not present any new theoretical contributions to machine learning in this work. However, for the existing theoretical claims and assumptions underlying the diffusion Schrödinger bridge (DSB) framework on which our method is based, we provide detailed proofs in the appendices for interested readers.

Similarly, this work does not introduce new astrophysical discoveries. The adopted simulation codes are based on well-established astrophysical theories and assumptions, with detailed descriptions and references also provided in the appendices.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: We fully disclose the implementation details of our key modeling components, training, and inference procedures in the main paper and appendices. We also provide those implementation code as part of the supplementary materials in our submission.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

- (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We have provided the core code as part of the submission, and we will release the full codebase and data upon paper acceptance.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: We have provided implementation details covering the necessary training and test details to understand the results.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: All the quantitative results reported in this paper include error bars that define appropriate information about the statistical significance of the experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: We have provided the information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: This research conforms with the NeurIPS Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: We have included a statement on broader impacts in the beginning of the appendices. However, we do not foresee any significant societal impact of our work as the main application is within astronomical data for star formation.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: This paper does not pose the described risks.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The code, data, and models used in this paper has been properly credited with corresponding license and terms of use.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.

- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [\[Yes\]](#)

Justification: We will release our data and codebase of this work.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [\[NA\]](#)

Justification: This research does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [\[NA\]](#)

Justification: This research does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. Declaration of LLM usage

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: We do not use LLMs for any core design of the proposed method, but only for minor writing editing.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.

Technical Appendices and Supplementary Material

A statement of broader impacts is included in Section A. Section B provides additional details on the relevant astrophysical background for readers interested in research around Giant Molecular Clouds (GMCs) and star formation. Further experimental details, including extended qualitative comparisons across different methods, are presented in Section C. We have also included our **core code** as part of the supplementary material, with the full codebase and datasets to be released upon acceptance.

A Statement of broader impacts

We apply the Diffusion Schrödinger Bridge framework to observational inverse prediction problems in astronomy. We do not anticipate any immediate societal risks or ethical concerns associated with this research. Our work focuses on advancing scientific understanding in a domain with minimal direct human impact, and all data used in this study are derived from publicly available astronomical observations or licensed simulations.

Beyond its immediate scientific goals, this research contributes positively to the broader advancement of interdisciplinary science. By demonstrating the application of a principled generative modeling framework in a physically grounded domain, our work helps bridge the gap between data-driven machine learning and theory-driven physical modeling. This approach not only promotes interpretability and robustness in ML research but also supports the long-term vision of using AI to accelerate scientific discovery. We hope this study encourages further collaboration between the machine learning and astrophysics communities, especially in leveraging modern ML techniques to understand complex natural systems.

B Astrophysical background

We provide extended details about GMCs and star formation research for readers interested in the astrophysical background related to this work. In this work, we focus on studying the physical conditions and processes within giant molecular clouds (GMCs), by inferring the precise distributions of various critical physical quantities from limited observational data.

Giant Molecular Clouds (GMCs) are key structures within the interstellar medium (ISM) of galaxies, composed primarily of cold gas and dust. The ISM, which fills the space between stars, also contains cosmic rays and magnetic fields and plays a central role in galactic evolution by regulating star formation rates (Spitzer Jr, 2008). GMCs are of particular interest because they harbor most of the dense gas in the ISM, which is essential for the formation of stars (McKee and Ostriker, 2007). The physical conditions within GMCs are highly variable and complex, exhibiting spatial and temporal variations in density, temperature, velocity, and magnetic field strength and orientation. These factors give rise to a wide array of astrophysical phenomena, including the formation of protostars (Heyer and Dame, 2015), star clusters (Krumholz et al., 2019), and complex organic molecules (Bisbas et al., 2023). Among all physical quantities, we study two fundamental quantities in this work, which are the gas density (ρ) and the magnetic field (B).

B.1 Gas density

The gas density (i.e., mass per unit volume), often measured as the number density of hydrogen nuclei (H), is considered one of the most fundamental quantities in GMC studies. Under the assumption of an abundance of one He nucleus for every 10 H nuclei in interstellar gas, we have a mass per H nucleus of $\mu_H = 1.4m_H = 2.34 \times 10^{-24}g$, leading to $n_H = 1cm^{-3}$ equivalent to $\rho = 2.34 \times 10^{-24}g cm^{-3}$.

The gas density directly governs gravitational collapse, sets the local timescale for star formation, and strongly influences the thermal and chemical evolution of the gas. For example, denser regions within GMCs are more likely to become gravitationally unstable, leading to the formation of stars or star clusters. Moreover, the density controls the rates of radiative cooling and chemical reactions, such as molecule formation and dust grain processes, which are crucial for understanding the lifecycle of interstellar gas and the conditions for planet formation. Accurately inferring the spatial distribution of density within GMCs is therefore essential for modeling the early stages of stellar evolution.

In astronomical studies, it is usually very challenging to precisely quantify the number density of GMCs from observations. The traditional astrostatistical approach relies on the observations of column density and certain assumptions on the geometry of the clouds. For example, a cylindrical geometry for filamentary structures or spherical geometry for dense cores (Palmeirim et al., 2013). Bisbas et al. (2021, 2023) later propose an empirical power law to convert the observed column density to the mean number density of GMCs based on MHD simulations (Wu et al., 2015), which has also been adapted in our work as two baseline methods (i.e., two-component power-law conversion and three-component power-law conversion). Despite being effective in similar simulation data, the prediction performance of these methods often suffers from a significant drop when new observations present a large distributional shift compared to the original fitting data (i.e., OOD testing cases). In our work, we aim to introduce a machine learning based approach with better robustness and generalization ability.

B.2 Magnetic field

Magnetic fields, often measured in microgauss (μG), represent another critical physical quantity within the interstellar medium (ISM), playing a fundamental role in shaping the structure and regulating the dynamics of GMCs (Crutcher, 1999, 2012). They can provide support against gravitational collapse, guide anisotropic gas flows, and influence the morphology of star-forming regions through magnetic tension and pressure. In combination with turbulence, magnetic fields are believed to significantly affect the efficiency and fragmentation of star formation across spatial scales.

Similar to the previous density estimation, estimating the magnetic field strength within GMCs is a challenging task. The Davis–Chandrasekhar–Fermi (DCF) method is a widely adopted technique in astrophysics to estimate the plane-of-sky magnetic field strength B_{POS} in molecular clouds based on observed dust polarization (Davis, 1951; Chandrasekhar and Fermi, 1953; Beck, 2016), which relates the magnetic field strength to the polarization angle and other observable properties. Its core assumption is that an equipartition exists between the magnetic field energy and the turbulent kinetic energy of the gas. Specifically, the DCF method states that the relation between the gas density ρ , the nonthermal velocity dispersion σ_V , and the polarization angle dispersion σ_{PA} gives the POS magnetic field strength B_{POS} as:

$$B_{POS} = f \sqrt{4\pi\rho} \frac{\sigma_V}{\sigma_{PA}}, \quad (7)$$

where f is a correction factor accounting for projection and geometric effects. More recently, a variant of DCF method (Skalidis and Tassis, 2021) takes the compressible modes into account and modifies the previous Eq. 7 to:

$$B_{POS} = \sqrt{2\pi\rho} \frac{\sigma_V}{\sqrt{\sigma_{PA}}}. \quad (8)$$

While widely used, this method suffers from several limitations: (1) the equipartition assumption may not hold in strongly magnetized or shock-dominated regions; (2) σ_ϕ becomes unreliable in regions with low signal-to-noise ratio or complex field morphology; and (3) accurate density and turbulence estimates are required for robust magnetic field inference, which are themselves uncertain. Moreover, DCF provides only an average value over large regions and lacks the ability to resolve fine-scale spatial structures.

In this work, we aim to bypass the limitations of traditional analytic approximations by leveraging machine learning to directly infer the underlying magnetic field distribution from observables, jointly with other key physical quantities such as gas density. Our proposed framework thus offers a more flexible and data-driven solution, especially under conditions with limited observability or strong distribution shifts.

In Fig. 5, we visualize an example of molecular cloud data used in machine learning methods from Xu et al. (2025), including the map of column density with magnetic field directions, the magnetic field angle dispersion, the LOS velocity dispersion, and the projected magnetic field strength.

C Additional experimental results and analysis

We provide extended details about the experiments in this section.

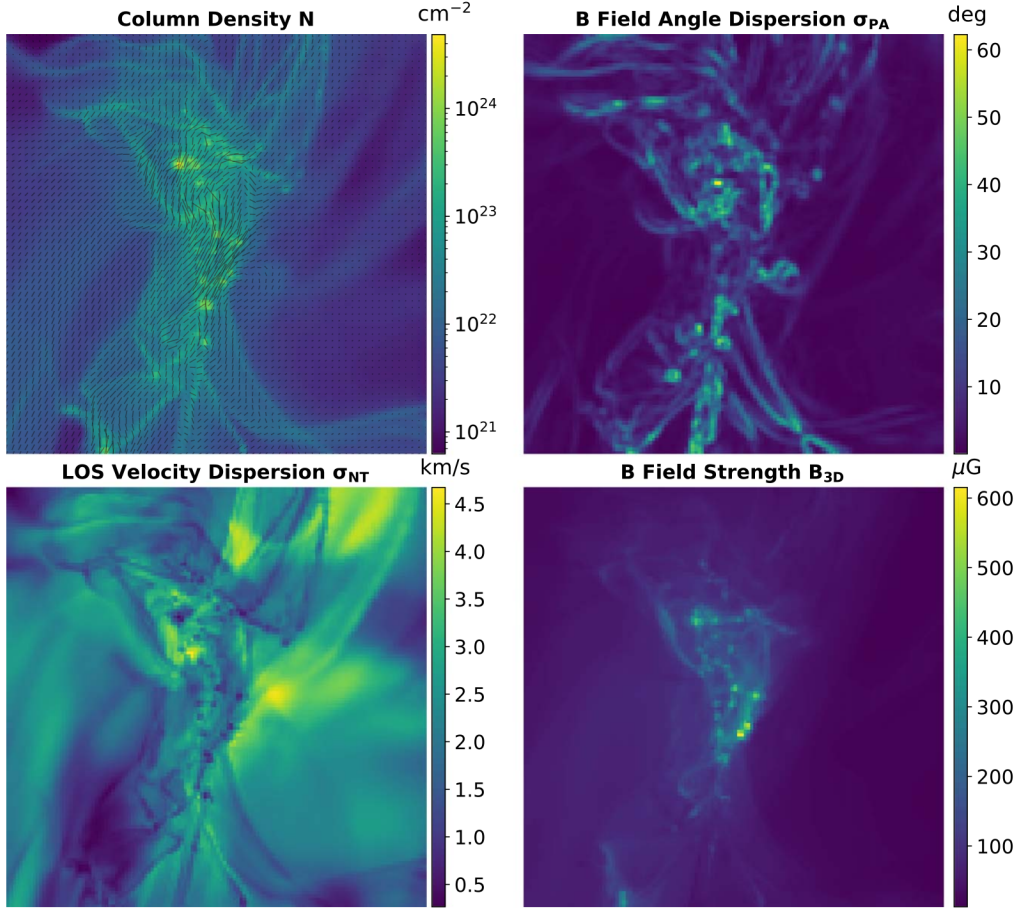


Figure 5: **Illustration of different synthetic observables from (Xu et al., 2025) for density and magnetic field estimation.** Top left: map of column density with magnetic field directions; Top right: magnetic field angle dispersion; Bottom left: LOS velocity dispersion; Bottom right: projected magnetic field strength.

C.1 More details about the magnetohydrodynamics data simulations

We follow the same simulation setups used in prior works that apply conditional DDPMs to astrophysical observational prediction (Xu et al., 2023b, 2025; Wu et al., 2020; Hsu et al., 2023). While the design and analysis of physical simulations represent an important research direction on their own that is distinct from the machine learning focus of this work, we provide only the essential background here and refer interested readers to the original astrophysical literature for further details.

Density simulation. For the density simulations, we adopt the magnetohydrodynamic (MHD) simulations of colliding and non-colliding GMCs following prior works (Wu et al., 2020; Hsu et al., 2023), performed using the adaptive mesh refinement (AMR) code Enzo (Bryan et al., 2014). The simulations include self-gravity, magnetic fields, and thermal processes under a photodissociation radiation field ($G_0 = 4$ Habings) and cosmic-ray ionization rate $\zeta = 10^{-16} \text{ s}^{-1}$. Two clouds with radii of 20 pc are initialized in a 128^3 pc^3 domain with a 256^3 grid resolution.

Each cloud starts with $n_H = 83 \text{ cm}^{-3}$ and $T = 15 \text{ K}$, while the ambient gas is set at $n_H = 8.3 \text{ cm}^{-3}$ and $T = 150 \text{ K}$ to maintain pressure balance. A multiphase temperature structure develops over time, with high-density regions ($n_H \gtrsim 10^3 \text{ cm}^{-3}$) cooling to 10–20 K, and low-density regions ($n_H \lesssim 10 \text{ cm}^{-3}$) reaching up to 1000 K. Magnetic fields are initialized at 10–50 μG , oriented 60° to the collision axis. The refinement strategy ensures that the local Jeans length is resolved by at least eight cells.

The simulation runs span 5 Myr, with snapshots taken at 3 and 4 Myr. Column density maps and corresponding line-of-sight (LOS) mass-weighted number density maps are extracted at multiple physical scales (32, 16, 8, and 4 pc), and projected to 128×128 pixel images. The resulting dataset contains 7179 samples, with an approximate 80/20 split between training and testing. For more details, please refer to (Wu et al., 2020; Hsu et al., 2023).

Magnetic field simulation. For magnetic field prediction tasks, we adopt ideal magnetohydrodynamics (MHD) simulations following the setup in prior work (Wu et al., 2020; Hsu et al., 2023), using the MUSCL-Dedner method and HLLD Riemann solver within the AMR code ENZO (Bryan et al., 2014). These simulations incorporate self-gravity, magnetic fields, and thermal processes under a photodissociation radiation field with $G_0 = 4$ and a cosmic-ray ionization rate of $\zeta = 10^{-16} \text{ s}^{-1}$. Each simulation initializes two clouds (radius 20 pc) in a 128 pc^3 domain at 256^3 resolution, with initial hydrogen number density $n_{\text{H}} = 83 \text{ cm}^{-3}$, temperature $T = 15 \text{ K}$, and solenoidal turbulence satisfying $v_k^2 \propto k^{-4}$ for $2 \leq k \leq 20$. The magnetic field is initialized at a 60° angle to the collision axis with strengths of 10, 30, and $50 \mu\text{G}$ across different cases, and four additional refinement levels are used to resolve the local Jeans length. For each magnetic field configuration, we model both colliding and non-colliding GMC scenarios, where colliding clouds are offset by $0.5 R_{\text{GMC}}$ and have a relative velocity of 10 km/s . These simulations run up to 4.1 Myr and capture early evolutionary phases before star formation begins.

To generate realistic observables, we compute maps of column density, line-of-sight (LOS) mass-weighted polarization angle, LOS nonthermal velocity dispersion, and projected 3D magnetic field strength across varying AMR levels. The final dataset consists of 25,479 images at 128×128 resolution, sampled across physical scales of 32, 16, 8, and 4 pc. To enhance diversity, we include multiple projection angles and aggregate LOS effects without manually disentangling contributions such as turbulence or cloud-scale motions, consistent with observational limitations.

C.2 Additional training details

As we have mentioned in the main paper, our proposed *Astro-DSB* benefits from a faster convergence speed and thus less computational resources compared to conditional DDPMs and vanilla DSB models. In Fig. 6, we compare the training curves among different variants of DSB as additional ablation analysis.

Compared to previous works that leverage conditional DDPMs as the prediction framework (Xu et al., 2023a, 2025), our *Astro-DSB* converges around 25k iterations as shown in Fig. 6(d), which corresponds to approximately 10 GPU hours. In contrast, conditional DDPMs take around 40 GPU hours to train.

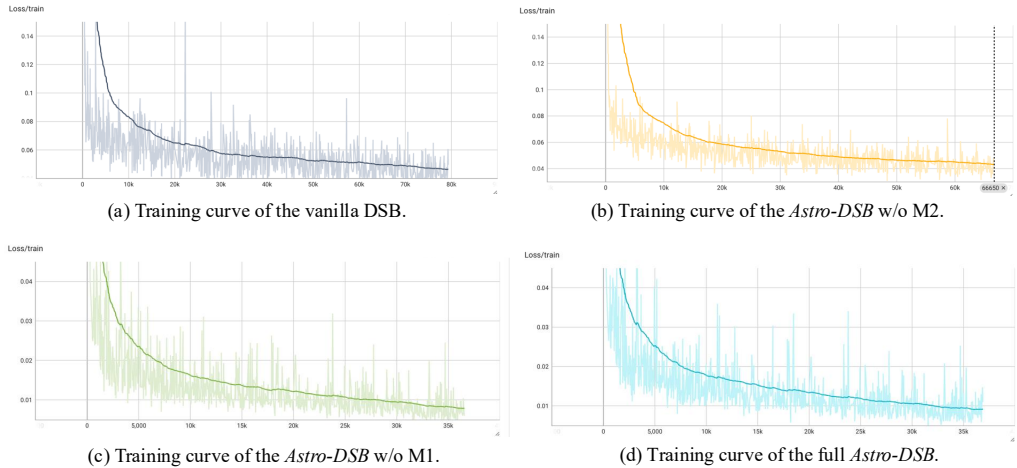


Figure 6: **Training curve comparison of different variants of DSB for ablation analysis.** We show that the proposed M2 component on the observable enhancement contributes to a faster training convergence speed. Best viewed in color and with zoom-in.

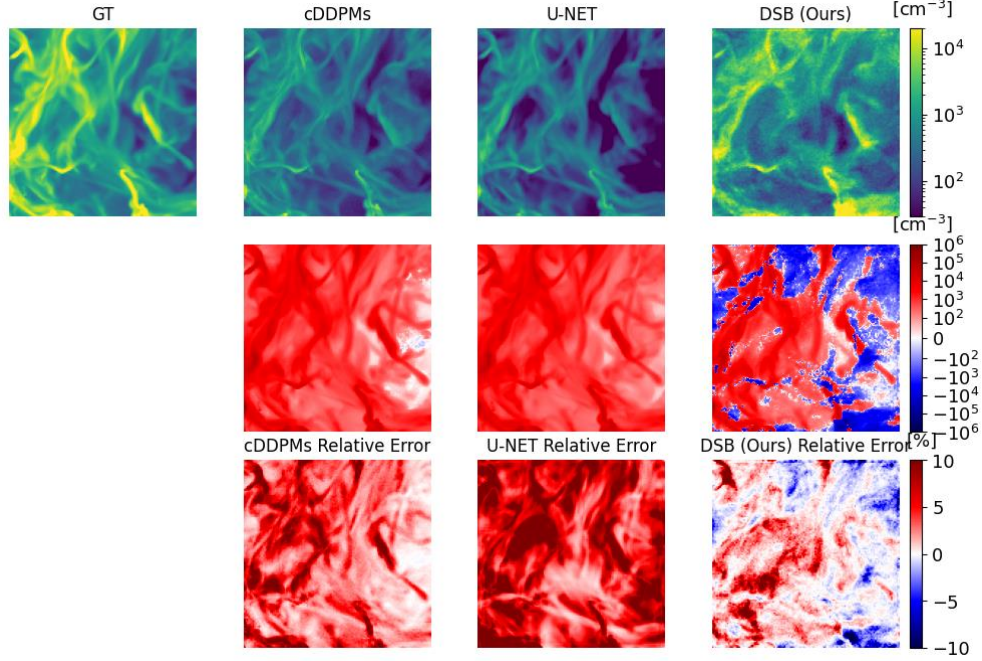


Figure 7: **Additional visualization of relative error patterns from different machine learning methods for the density prediction task on OOD testing cases.** In the first row, we qualitatively show the GT and prediction results from conditional DDPMs, U-Net, and our proposed *Astro-DSB* method. The second and third rows present the relative errors computed by “GT - pred” in absolute values and the values after log transformation, respectively. Compared to other ML methods that tend to underpredict in OOD testing cases, our *Astro-DSB* achieves less biased and more robust prediction performance.

C.3 More details about experiments on simulation data

As explained in the main paper, evaluations in astrophysical prediction particularly emphasize the relative error patterns exhibited between the prediction results and the ground-truth distributions. In Fig. 7, we present additional visualization of relative error patterns from different machine learning methods for the density prediction task on OOD testing cases. The first row shows the ground truth and prediction results from conditional DDPMs, U-Net, and our proposed *Astro-DSB*. The second and third rows visualize the corresponding relative errors, computed as “GT - Pred” in absolute values and after applying log transformation, respectively. Compared to DDPMs and U-Net, which tend to significantly underpredict dense structures in the OOD regime, *Astro-DSB* produces more balanced and spatially consistent residuals. These visual results further support the claim that our method achieves improved robustness and lower bias when generalizing beyond training distributions.

To better understand the impact of each design component, we compare the distribution of relative errors for different methods in Fig. 8. Traditional power-law baselines (p2, p3) and vanilla DDPMs (cdm) exhibit large biases and heavy tails, especially under distribution shifts. The vanilla DSB (psbv1) slightly reduces error spread, while our proposed components—M1 and M2—each contribute to sharpening and centering the PDF. The full model (psbv4) demonstrates the narrowest and most symmetric error distribution, highlighting its robustness and well-calibrated behavior in OOD settings.

C.4 More details about experiments on real observations

We further compare our proposed *Astro-DSB* with previous machine learning methods on real observational data from the Taurus B213 region, focusing on the inferred number density distribution. As shown in Fig. 4, the prediction of *Astro-DSB* closely matches the spatial structures and filament morphology present in the original Herschel column density map. In contrast, the prediction from

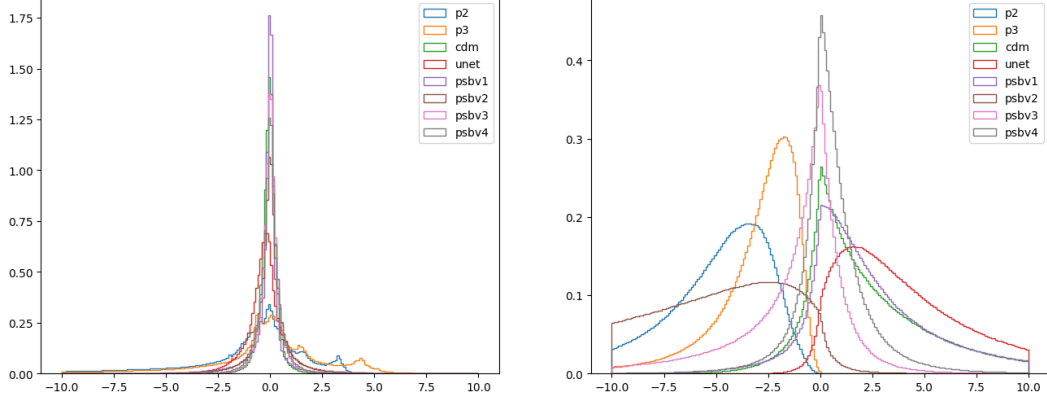


Figure 8: **PDF comparisons across different methods for the density prediction task on ID and OOD data.** Methods $p2$ and $p3$ correspond to two- and three-component power-law conversions, respectively. cdm denotes conditional DDPMs. $psbv1$ is the vanilla diffusion Schrödinger bridge without our proposed modules. $psbv2$ and $psbv3$ remove M1 (noise alignment) and M2 (observable enhancement), respectively, while $psbv4$ represents the full *Astro-DSB* model.

CASI-2D (Xu et al., 2020a) fails to capture fine-scale density variation and underestimates extended dense regions. The conditional DDPMs produce more coherent predictions but still smooth out localized dense structures. Overall, *Astro-DSB* exhibits better spatial continuity, structure recovery, and dynamic range, demonstrating its superior robustness and generalization capability on real-world astrophysical observations.

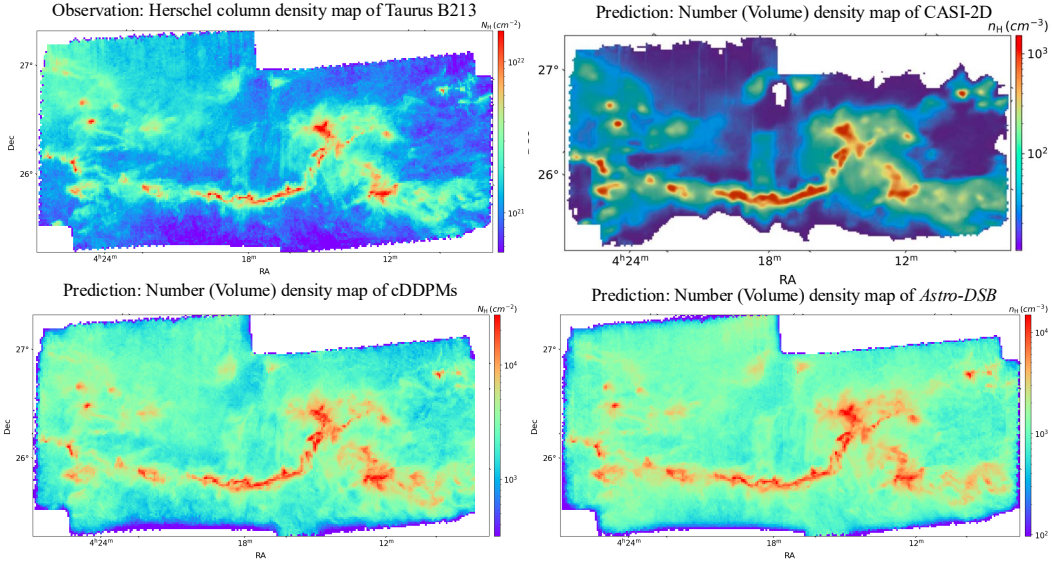


Figure 9: **Additional qualitative comparisons among different machine learning methods on Taurus B213 number density prediction.** CASI-2D (Xu et al., 2020a) is a previous ML method with U-Net architecture trained with discriminative reconstruction loss.