
Beyond Equilibrium: Non-Equilibrium Foundations Should Underpin Generative Processes in Complex Dynamical Systems

Anonymous Author(s)

Affiliation

Address

email

Abstract

1 This position paper argues that next-generation non-equilibrium-inspired generative
2 models will provide the essential foundation for better modeling real-world complex
3 dynamical systems. While many classical generative algorithms draw inspiration
4 from equilibrium physics, they are fundamentally limited in representing systems
5 with transient, irreversible, or far-from-equilibrium behavior. We show that non-
6 equilibrium frameworks naturally capture non-equilibrium processes and evolving
7 distributions. Through empirical experiments on a dynamic Printz potential system,
8 we demonstrate that non-equilibrium generative models better track temporal
9 evolution and adapt to non-stationary landscapes. We further highlight future
10 directions such as integrating non-equilibrium principles with generative AI to
11 simulate rare events, inferring underlying mechanisms, and representing multi-scale
12 dynamics across scientific domains. Our position is that embracing non-equilibrium
13 physics is not merely beneficial—but necessary—for generative AI to serve as a
14 scientific modeling tool, offering new capabilities for simulating, understanding,
15 and controlling complex systems.

1 Introduction

17 Since Boltzmann introduced statistical mechanics in the 1870s, a groundbreaking framework in
18 mathematical physics, the core philosophy of this theory has remained consistent: uncovering the
19 statistical laws governing the motion of microscopic particles, which initially appear to move in an
20 irregular manner. In the 1950s, Jaynes [1] extended this theory to information theory and statistical
21 learning, proposing that if the motion of microscopic particles is viewed as data, this data must
22 adhere to similar mathematical principles as statistical mechanics, from Boltzmann distribution [2, 3]
23 to Energy-based models (EBMs) [4–11], from variational free energy theory [3, 12] to Variational
24 Encoders (VAEs) [13–17], from Schrödinger bridges [18–20] to the time-reversal duality in the
25 Diffusion model [21–23]. In the 1980s, a series of groundbreaking discoveries [4–7, 24] in artificial
26 intelligence emerged, many of which were inspired by the principles of statistical mechanics. After
27 a period of relative dormancy, the intersection of artificial intelligence and statistical physics has
28 recently sparked renewed interest and innovation [20–23, 25–30].

29 Although generative models in machine learning are typically designed for practical tasks such
30 as image, text, and audio synthesis, they share a core philosophical foundation with statistical
31 physics: both seek to model and generate probability distributions from high-dimensional, noisy
32 data. However, many real-world scientific and physical systems are inherently dynamical, exhibiting
33 time-dependent, path-dependent, and often irreversible behavior. Capturing such complex temporal
34 evolution remains a fundamental challenge in generative modeling, as most existing frameworks
35 assume static data distributions or stationarity. This gap motivates a deeper integration between

generative modeling and statistical physics, especially in the context of dynamical systems that operate far from equilibrium. In this position paper, we argue that generative models grounded in non-equilibrium statistical mechanics provide a principled and expressive framework to simulate, predict, and understand the temporal evolution of such systems, going beyond the limitations of equilibrium-based models.

In Section 2, we introduce the connection between statistical physics and generative modeling from three aspects: equilibrium-inspired models, non-equilibrium-inspired models, and non-physics-inspired models. In Section 2.1, we show how equilibrium statistical mechanics inspires machine learning. Equilibrium systems follow Boltzmann distributions, linking state probabilities to energy. Starting with the Ising model, we introduce energy-based models (EBMs) such as Hopfield networks [4] and RBMs [5–7], which use energy functions to model data. Modern EBMs scale this approach to high-dimensional data using deep networks and Langevin dynamics [9–11]. Despite their power, EBMs face challenges including intractable normalization, misalignment with non-equilibrium data, and poor modeling of dynamics. Section 2.2 outlines how non-equilibrium physics reframes generative modeling as distribution transformation via Markov chains [19, 31, 32]. Diffusion models [25, 23] implement this through stochastic dynamics. Advances such as Fokker-Planck constraints [26], Poisson flows [27], and Schrödinger bridges [20, 28] improve sampling efficiency. In Section 2.3, we relate flow-based models [33–35], VAEs [13, 14, 16], autoregressive models [36–38], and GANs [39] to statistical physics, noting their limitations in capturing system dynamics.

In Section 3, we empirically demonstrate that non-equilibrium generative models are better suited for such systems, as they capture time-asymmetric dynamics and transient steady states more effectively. Using a modified Printz potential system with time-varying energy fields governed by overdamped Langevin dynamics, we compare two generative strategies: (i) equilibrium-based Boltzmann sampling and (ii) non-equilibrium sampling via denoising Langevin dynamics. Experiments show that the non-equilibrium approach consistently achieves lower Jensen-Shannon divergence when generating time-dependent particle distributions, better reflecting the evolving energy landscape. Unlike equilibrium methods, which rely on static distributions, non-equilibrium models adaptively track gradient flows, enabling more accurate simulation of dynamic transitions. These results underscore the necessity of non-equilibrium frameworks for modeling the temporal evolution of open, evolving systems.

In Section 4, we further discuss the combination of non-equilibrium statistical mechanics and generative AI from two aspects: (1) non-equilibrium-inspired AI models for physical science and complex dynamical systems; and (2) non-equilibrium physics for more advanced generative models.

This position paper argues that non-equilibrium physics-inspired generative algorithms represent a new frontier for modeling dynamical systems in AI for science. Dynamical systems, ranging from molecular reactions and fluid turbulence to biological networks and climate dynamics, are inherently governed by time-asymmetric, dissipative, and stochastic processes far from equilibrium. Non-equilibrium-inspired algorithms treat data as outcomes of underlying time-evolving processes, offering a principled framework to learn, simulate, and control complex system trajectories. By integrating physical priors such as entropy production, irreversible flows, and fluctuation theorems, these models have potential to more faithfully model the real-world dynamics across scientific domains.

2 Revisiting Generative Models Through Statistical Mechanics

2.1 Energy-based generative models inspired by equilibrium statistical mechanics

Equilibrium describes a macroscopically stable state, where the Boltzmann distribution [40] relates state probability to energy:

$$P_i = \frac{1}{Z} e^{-\frac{E_i}{kT}}, \quad (1)$$

with E_i as the energy of state i , T the temperature, and Z the partition function. This allows energy-based modeling of otherwise intractable distributions.

Building on this, energy-based generative models [4, 41, 10] approximate data distributions via energy functions. They root in the Ising model [42], where discrete spin states interact with neighbors:

$$E(\mathbf{s}) = - \sum_{\langle ij \rangle} J_{ij} s_i s_j, \quad (2)$$

85 with $J_{ij} > 0$ lowering energy when spins align.

86 Hopfield networks [4] generalize this to neural systems, encoding memory through Hebbian learning
87 with energy

$$E(\mathbf{s}) = -\frac{1}{2} \sum_{i \neq j} W_{ij} s_i s_j - \sum_i \theta_i s_i. \quad (3)$$

88 RBMs [41] further extend this to bipartite graphs with visible $\mathbf{v}$ and hidden $\mathbf{h}$ layers:

$$E(\mathbf{v}, \mathbf{h}) = - \sum_i b_i v_i - \sum_j c_j h_j - \sum_{i,j} W_{ij} v_i h_j, \quad (4)$$

89 learning data distributions via contrastive divergence [8]. Together, these models exemplify how
90 equilibrium statistical mechanics informs generative learning.

91 Modern energy-based models combine equilibrium statistical mechanics with deep networks and
92 Langevin dynamics to scale to high-dimensional, multi-modal data with flexible composition and
93 sampling [9–11, 43, 44]. However, equilibrium-inspired EBM face three major limitations. First,
94 the partition function Z is intractable, requiring approximations like contrastive divergence [8] or
95 MCMC [45–47], which can yield unstable training. Second, their reliance on stationary Boltzmann
96 assumptions mismatches the non-equilibrium, driven nature of real-world processes. Third, they lack
97 inherent mechanisms for modeling temporal transitions, limiting applicability to dynamic systems.

98 2.2 Generative models inspired by non-equilibrium statistical mechanics

99 While in theory any target distribution $P(\mathbf{x}) = \frac{\phi(\mathbf{x})}{Z}$ can
100 be modeled via a flexible neural network ϕ , in practice,
101 computing the partition function Z and sampling from
102 $P(\mathbf{x})$ are computationally costly. Non-equilibrium statisti-
103 cal physics [19, 52, 53, 31, 54, 32, 55, 56] addresses this
104 by constructing a Markov chain that transforms a simple
105 distribution (e.g., Gaussian) into the target via diffusion.
106 This shifts the learning objective from fitting $P(\mathbf{x})$ directly
107 to modeling diffusion dynamics. By breaking the problem
108 into small local perturbations, this paradigm [21] offers a
109 more tractable alternative to directly specifying a global
110 potential.

Table 1: Generative models inspired by non-equilibrium statistical mechanics.

Algorithms	Drift-term	Diffuse-term
DDPM [22]		✓
SMLD [25]		✓
VDM [16]		✓
SDEs [23]	✓	✓
RDM [48]	✓	✓
Flow++ [49]	✓	
PFGM [27]	✓	
DiffFlow [50]	✓	✓
DSB [20]	✓	✓
DSBM [28]	✓	✓
LightSB-M [51]	✓	✓

111 Inspired by non-equilibrium statistical physics, diffusion models systematically and gradually destroy
112 the data distribution structure through an iterative forward diffusion process. The model then
113 learns the reverse diffusion process to recover the data structure [22, 25]. Once the score function
114 $s_\theta(\mathbf{x}) = \frac{\partial \log P(\mathbf{x})}{\partial \mathbf{x}}$ is trained, Langevin dynamics $\mathbf{x}_{i+1} = \mathbf{x}_i + \epsilon \frac{\partial \log P(\mathbf{x})}{\partial \mathbf{x}} + \sqrt{2\epsilon} \mathbf{z}_i$ [24, 57] are
115 iteratively applied to sample from the distribution, where $\mathbf{z}_i \sim \mathcal{N}(0, I)$. Song et al. [23] further
116 unified the continuous-time evolution between the data distribution and the prior Gaussian distribution
117 into a set of forward and backward stochastic differential equations (SDEs) (see Appendix A)

$$\begin{cases} d\mathbf{x}_t = f(\mathbf{x}_t, t)dt + \sigma(t)dW_t, \\ d\mathbf{x}_t = \left[f(\mathbf{x}_t, t) - \sigma^2(t) \frac{\partial \log P(\mathbf{x}, t)}{\partial \mathbf{x}} \right] dt + \sigma(t)d\tilde{W}_t, \end{cases} \quad (5)$$

118 where the choice of drift coefficient $f(\mathbf{x}, t)$ defines different SDEs, and W_t is the standard Wiener
119 process scaled by the diffusion coefficient $\sigma(t)$.

120 The unified differential equation form of diffusion models has inspired further integration of non-
121 equilibrium statistical physics. From the perspective of stochastic processes, the forward process
122 described in Equation 5, which represents the spatiotemporal evolution of data density, can be
123 characterized by the classical Fokker-Planck equation (FPE) [58]. Therefore, in principle, the data
124 density at all time steps can be recovered by solving the FPE, requiring only guidance of the noise
125 density without additional learning [59]. However, in the work of Lai et al. [26], they found that
126 the score function optimized through score matching [60, 61] fails to satisfy the density distribution
127 derived from the FPE spontaneously. By incorporating a regularization term loss guided by the FPE
128 as a physical prior, diffusion models can provide more accurate density estimates for synthetic data.

To reduce the computational burden of classical diffusion processes, recent work explores more efficient diffusion paths. Xu et al.[27] proposed Poisson flow generative models, initializing from a uniform distribution on a high-dimensional hemisphere and targeting a data distribution on $\mathbf{z} = 0$, analogous to electric field lines. Beyond physics-inspired designs[62], Lipman et al.[63] introduced conditional flow matching, a generalized framework for learning optimal transport paths between distributions. The Schrödinger bridge (SB)[18, 64, 65, 51], as an entropy-regularized optimal transport on path space, enables more efficient sampling than Langevin dynamics. Doucet et al.[20, 28] solved the SB via Iterative Proportional and Markovian Fitting, and showed that it can recover classical diffusion behaviors[25].

2.3 The connection of non-physics-inspired generative models to statistical physics

Flow-based models [33–35] construct invertible, differentiable maps from simple latent variables to complex data, enabling exact likelihoods and efficient sampling via the change-of-variables formula: $P(x) = P(z) |\det(\partial f^{-1}/\partial x)|$. Though not explicitly physics-inspired, flow models align with non-equilibrium theory. Their governing ODEs reduce to the continuity equation (see Appendix B):

$$\frac{\partial P(x, t)}{\partial t} = -\frac{\partial v P(x, t)}{\partial x}, \quad (6)$$

where v denotes system velocity. If $v = f - \partial \log P(x, t)/\partial x$, this becomes the Fokker–Planck equation with drift $f = v + \partial \log P(x, t)/\partial x$. Hence, flow models recover either Hamiltonian or dissipative dynamics, bridging generative modeling with non-equilibrium statistical physics.

Variational Autoencoders (VAEs) constitute a powerful class of generative models that leverage latent variables to represent high-dimensional data distributions [13–16, 66]. By introducing latent variables z with a prior distribution $P(z)$, VAEs define a probabilistic generative process $P_\theta(x|z)$ parameterized by neural networks. Due to the intractability of the marginal likelihood $P_\theta(x)$, VAEs approximate the posterior with a variational distribution $Q_\phi(z|x)$, and optimize the evidence lower bound (ELBO) [13, 67]:

$$\log P_\theta(x) \geq \mathbb{E}_{Q_\phi(z|x)}[\log P_\theta(x|z)] - D_{KL}(Q_\phi(z|x)||P(z)). \quad (7)$$

Next we will show that the ELBO takes the same mathematical form of variational free energy F in equilibrium systems [68], which satisfies $F = -kT \ln Z$, where $Z = \int \exp[-E(z)/kT] dz$ is the partition function of Boltzmann distribution in Equation 1. However, the real Boltzmann distribution is hard to obtain; hence, we approximate it by a variational distribution $Q(z)$. Then, we can define the variational free energy of the equilibrium system

$$F(Q) = \mathbb{E}_Q(z)[E(z)] - kTH(Q) \geq F, \quad (8)$$

where $H(Q)$ is the variational entropy of the system. Compared with Equation 7 with 8, the ELBO parallels variational free energy, with its likelihood and KL terms corresponding to expected energy and entropy, thereby linking VAEs to both equilibrium and non-equilibrium statistical mechanics [16].

Autoregressive models. Unlike flow-based model and VAE, autoregressive models have little relation with physics [36, 69, 37, 38, 70]. Their main idea is to decompose high-dimensional data distributions into products of conditional probabilities using the chain rule:

$$P(\mathbf{x}) = \prod_{i=1}^n P(x_i|x_{<i}). \quad (9)$$

Autoregressive models have achieved great success in text and image generation [36, 69, 37, 38, 70–73]. Autoregressive models have shown promise in video generation [74], but often produce artifacts that violate physical laws due to their lack of explicit temporal dynamics or causal structure. Unlike physics-informed models (e.g., Langevin dynamics or neural ODEs/SDEs), they capture statistical correlations without modeling system evolution, making them insufficient for representing real-world dynamical processes without additional structural constraints.

Generative Adversarial Networks (GANs)[39] synthesize high-quality samples by training a generator and discriminator adversarially, bypassing explicit likelihood modeling. While successful in image generation[75–78], GANs lack mechanisms to model temporal evolution or causality. Since samples are independently drawn from latent space, they fail to capture time-dependent consistency or state transitions, making them unsuitable for dynamic tasks like video synthesis or trajectory prediction without architectural or hybrid modifications.

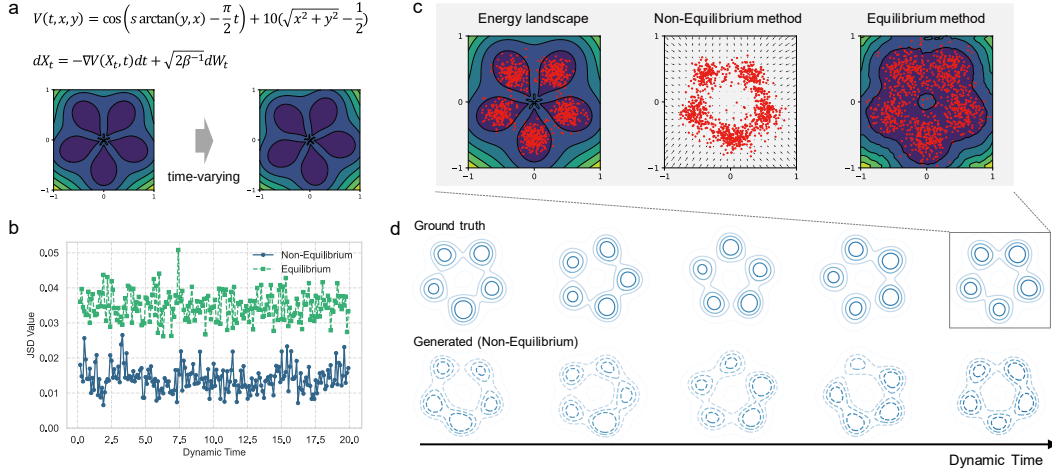


Figure 1: Equilibrium vs. non-equilibrium generation on a time-varying potential system. (a) Potential & dynamics equations. (b) Generation error comparison. (c) True energy landscape (left), non-equilibrium gradient field (mid), equilibrium energy field (right). (d) Generated distribution by non-equilibrium method.

3 Non-Equilibrium Generative Modeling Better Captures Dynamics in Evolving Systems

In this section, we demonstrate two viewpoints through experiments: (1) non-equilibrium dynamics are better suited for modeling systems without well-defined energy landscapes, such as those that remain far from equilibrium or exhibit only transient steady states; (2) non-equilibrium generative models more effectively capture dynamic behaviors and temporal evolution, aligning with the intrinsic time-asymmetry of many real-world systems.

To directly compare the differences in generative models guided by the principles of equilibrium and non-equilibrium processes for generative modeling of complex systems, we modify and conduct tests on a 2-dimensional Printz potential system (Figure 1a) in relevant literature [79, 80]. The particle’s motion follows overdamped Langevin dynamics, with its drag term described by the potential energy landscape. The governing equation of potential function involves a time-dependent term, which leads to a time-varying energy landscape (rotating). As the potential energy changes over time, the particle’s motion is continuously perturbed, causing the system to remain in non-equilibrium transitions between different states. Our generative modeling objective is to learn the time-varying energy field $V(t, x, y)$ and subsequently generate reliable particle distributions $p(x, t)$. Inspired by the distinctions between equilibrium and non-equilibrium processes, we adopt two approaches for modeling the energy field: i) Equilibrium method, and ii) Non-equilibrium method. We simulate particle trajectories under 2,000 initial conditions for training and testing, using the Jensen-Shannon Divergence (JSD) of the generated particle distributions as the error metric. Detailed methods and settings can be found in Appendix D.

In generated time-varying particle distributions, the JSD of the non-equilibrium method consistently remains lower than that of the equilibrium method (Figure 1b), highlighting its strong generative modeling potential for complex systems. Figure 1c visualizes the true energy landscape at a dynamic moment and the predictions from both methods. The non-equilibrium method uses a gradient field to reflect the energy function’s influence on particles, sampling the true distribution through a constructed non-equilibrium process (denoising Langevin dynamics). In contrast, the equilibrium method simply predicts time-varying energy values and performs Boltzmann distribution sampling. Evolving open systems often exhibit time-varying energy landscapes, rarely achieving or only briefly maintaining equilibrium. Non-equilibrium methods do not rely on the system reaching equilibrium but dynamically track the evolution of energy gradients. By aligning the generative process with the inherent non-equilibrium characteristics of the data, non-equilibrium methods accurately model the static distribution at specific moments and successfully capture the system’s temporal evolution (Figure 1d). Code and data for reproduction are open source¹.

¹Anonymous repository.

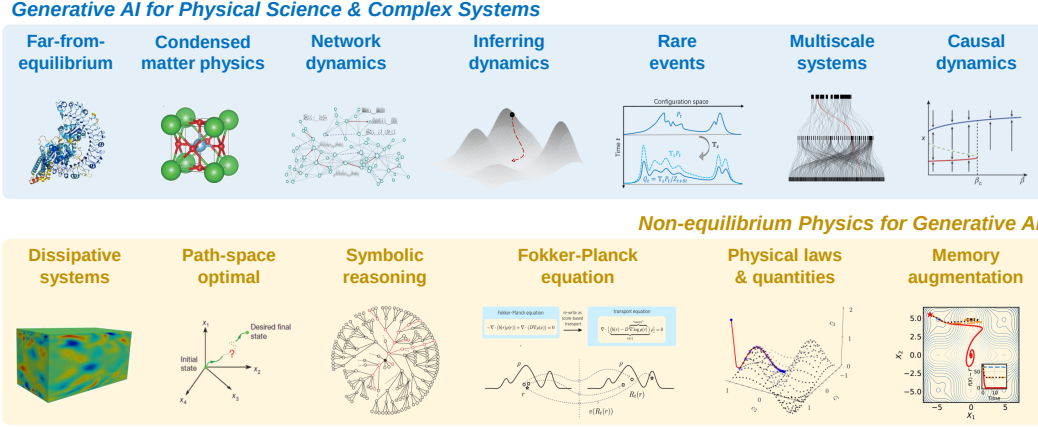


Figure 2: Outlook or future directions of both non-equilibrium physics and generative AI.

4 Outlook or future directions

In this section, we discuss opportunities and future directions for utilizing non-equilibrium statistical mechanics in generative AIs, mainly regarding two open questions: (i) In what fields will generative algorithms inspired by non-equilibrium physics have potential great applications? (ii) Can we utilize the wealth of non-equilibrium knowledge to develop more advanced generative models?

4.1 Potential future application to physical science and complex systems

Modeling far-from-equilibrium dynamics via non-equilibrium generative processes. A fundamental challenge in scientific modeling is capturing dynamics far from equilibrium [81–87], where systems exhibit transient, nonstationary, and strongly nonlinear behavior far from steady-state regimes. Most current generative frameworks inspired by non-equilibrium physics, such as Langevin dynamics and score-based diffusion models, implicitly assume convergence to a stationary distribution. However, many real-world systems, including climate extremes, biological differentiation, and socio-technical transitions, evolve in regimes where no steady state exists. Modeling such systems requires generative processes that track non-stationary distributions over time, adapt to evolving drift fields, and encode entropy production or phase transitions explicitly. Future work may extend generative stochastic dynamics to incorporate time-dependent control parameters or auxiliary slow-fast variables, enabling generative inference over non-equilibrated trajectories. Moreover, integrating Fokker–Planck solvers with neural sequence models could offer tractable and expressive tools. This direction opens a pathway toward generative modeling not merely of statistical distributions, but of the full unfolding of irreversible, history-dependent physical processes.

Non-equilibrium generative models for condensed matter physics. Diffusion generative models are proving effective in condensed matter physics for modeling high-dimensional, non-equilibrium systems. They enable unsupervised synthesis of microstructures—from crystalline and amorphous phases to defect networks like dislocations—by progressively denoising random noise into physically plausible states. In free-energy landscape sampling, diffusion processes efficiently explore multimodal surfaces, identifying equilibrium basins and transition states [79]. By conditioning on parameters such as temperature or stress, they simulate phase transitions and interface evolution, including nucleation and grain growth [86, 88–90]. In multiscale material design [91, 92], these models bridge coarse and atomistic descriptions, optimizing alloy compositions. They enhance Monte Carlo and molecular dynamics via learned proposal distributions in Metropolis–Hastings schemes, accelerating sampling for long-timescale processes like glass relaxation. Diffusion models also aid inverse problems and data fusion, reconstructing 3D atomic structures [93, 94] and revealing underlying forces. With improved computational resources, they are poised to advance materials discovery [95, 96] and non-equilibrium phase transition modeling.

Integrating non-equilibrium generative models with complex network dynamics. Combining non-equilibrium generative algorithms with complex network dynamics provides a powerful framework for modeling real systems [97, 98]. In this hybrid approach, a non-equilibrium generative model captures statistical laws over node and edge attributes, while network evolution is governed by

time-varying graph operators or neural dynamics (e.g., graph Neural ODEs). Coupling the two creates a feedback loop: generative samples reshape network topology, and the evolving network modulates generative dynamics. This enables realistic synthesis of temporal graphs with non-stationary structure and complex phenomena like community formation or cascading failures. Applications span epidemic forecasting [99], information diffusion, and resilient infrastructure modeling [100]. As stochastic graph flows and physics-informed graph learning advance [101], this integration will deepen our understanding of networked non-equilibrium phenomena [90].

Inferring dynamics via non-equilibrium generative models Non-equilibrium generative models have primarily focused on reproducing complex data distributions from dynamic systems. A promising frontier is inverting this goal: using generative processes to infer hidden dynamical mechanisms [102–104]. Models such as score-based SDEs, Fokker–Planck frameworks, and Schrödinger bridges encode interpretable quantities like drift fields, entropy production, and dissipation, linking data trajectories to physical insights. Parameterizing dynamics via physics-informed neural architectures enables principled inference of evolution laws. This supports causal discovery [105], intervention analysis [106, 107], and mechanism-aware simulation in non-stationary systems. In particular, these models may recover drivers of rare events or phase transitions where equilibrium methods fail. Developing robust, physically grounded inverse modeling tools presents both a challenge and an opportunity for integrating AI with non-equilibrium statistical physics.

Non-equilibrium generative AIs for rare events. Rare events in complex systems, such as extreme climate [108] or abrupt phase transitions [83, 86, 109–111], are hard to sample and predict due to their low probability and dynamic complexity. Non-equilibrium generative models can be adapted to generate trajectories biased toward rare-event regions by conditioning on boundary states, modifying drift fields, or minimizing entropy-regularized action costs. Incorporating non-equilibrium principles [54, 31, 56, 32, 112, 113, 42] further enhances physical fidelity and sampling efficiency. This enables not only targeted generation of rare transitions but also estimation of their probabilities and underlying mechanisms. Such approaches are well-suited for scientific applications where traditional Monte Carlo methods fail, offering a new generative framework for forecasting, simulating, and interpreting rare events in non-stationary environments [114, 115].

Non-equilibrium generative AI for multi-scale systems Multi-scale structures are ubiquitous in real-world physical systems [116], e.g., the Rayleigh–Bénard convection in the thermal fluid systems [117], the dense particle clusters surrounded by the dilute broth in the gas-solid two-phase flow [118, 119], and hierarchical folding structure of proteins [120]. Generative AIs excel at modeling multi-scale structures of real-world data through guidance loss incorporation [121], posterior reconstruction [122], or conditional modeling [123]. As a result, they provide a powerful framework for understanding and capturing the multi-scale features of complex physical phenomena. In particular, renormalization-group-inspired generative models [124] hierarchically decompose complex systems via block transformations and multi-scale latents for efficient synthesis. In quantum many-body physics, coarse latents capture collective modes while finer latents reconstruct entanglement networks [125]. Financial modeling uses scale-invariant flows to generate market regimes across time horizons. Epidemiology employs layered generators to simulate disease spread from global patterns down to local outbreaks [126]. Protein folding [127, 128] benefits from wavelet-based latent layers that first form secondary structures, then refine atomic contacts. Ecological networks [79] adopt recursive block flows to model interactions from ecosystem down to community scales.

Modeling causal dynamics through non-equilibrium-inspired generative process A promising future direction is to extend non-equilibrium generative models toward causal reasoning [19, 31, 56, 107, 106, 129–132], specifically, modeling interventions and counterfactuals in complex dynamical systems, which is similar to the idea of building a phase diagram in thermodynamics. Many physical and biological processes evolve under structured external influences, where interventions (e.g., perturbing a force field or blocking a biochemical reaction) induce non-trivial changes in system trajectories. While current generative models capture statistical correlations or average trajectories, they do not inherently simulate responses under hypothetical or counterfactual scenarios. Embedding causal structure into generative dynamics, by parameterizing intervention operators within stochastic differential equations, or learning structural-functional factorizations over time, could enable sampling from correlational distributions. Bridging score-based diffusion models with recent advances in causal dynamics, and incorporating thermodynamic constraints on entropy flow under intervention, may allow generative models what could have been under alternative dynamic laws. This direction opens new possibilities for scientific modeling, policy design, and decision-making in complex systems.

304 4.2 Non-equilibrium physics for more advanced generative models

305 **Langevin dynamics integrated non-equilibrium EBM generative model for dissipative systems.**

306 Unlike equilibrium EBMs that satisfy detailed balance and model data via a Boltzmann distribution
307 reflecting purely conservative interactions, non-equilibrium EBMs offer a principled bridge between
308 non-equilibrium statistical mechanics and generative AI by viewing data as states of driven, dissi-
309 pative systems. In this setting, the energy function encodes both conservative and non-conservative
310 forces, yielding a time-dependent model distribution $P_t(x) \propto e^{-E_t(x)}$. Training typically relies on
311 contrastive divergence or score matching, avoiding the need to compute the intractable partition func-
312 tion. Langevin dynamics serves as a natural tool for both sampling and learning, iteratively combining
313 gradient descent with noise to explore low-energy regions. Importantly, one can embed physical
314 priors—such as conservation laws (mass, momentum, energy) and symmetry constraints (e.g., trans-
315 lational, rotational)—into the architecture or dynamics [133–139], improving generalization and
316 interpretability. These models have been applied to learning turbulent flow under Navier–Stokes
317 dynamics [140] and reconstructing free-energy landscapes in complex reactions [141, 79]. Remaining
318 challenges include estimating high-dimensional time-varying partition functions, ensuring long-term
319 consistency, and reducing the computational cost of Langevin-based sampling.

320 **Non-equilibrium processes for path-space optimal generative models** A promising research
321 direction lies in modeling distributions on path space, the space of time-indexed trajectories, by
322 leveraging variational principles from non-equilibrium statistical mechanics. In particular, the
323 Schrödinger bridge problem [18–20] and entropy-regularized optimal transport [142, 143] offer a
324 theoretical foundation for learning the most probable evolution between prior and target distributions
325 under stochastic dynamics. Future generative frameworks may adopt dynamic score matching,
326 stochastic control, or time-dependent diffusion bridges to sample from such path-space distributions.
327 These methods can capture time-asymmetry, non-Markovianity, and transient structures in complex
328 systems. Moreover, parameterizing drift and diffusion fields in neural SDEs with physical constraints
329 can enable controllable and interpretable generation of trajectories. Path-space generative modeling
330 thus opens a new frontier in AI for science, with applications in rare-event simulation, path transition,
331 and dynamical inference in multiscale systems [144]. Developing solvers for such high-dimensional,
332 temporal generative problems remains a core challenge and opportunity.

333 **Bridging non-equilibrium generative models and symbolic reasoning for scientific discovery.**

334 Non-equilibrium generative models have demonstrated remarkable capacity to simulate complex
335 dynamics in high-dimensional systems. Yet, their learned representations often remain opaque,
336 hindering interpretability and scientific generalization. A promising direction is to integrate these
337 models with symbolic reasoning frameworks [145]. Specifically, non-equilibrium models can generate
338 rich trajectory data or latent fields that reflect underlying dynamics. These outputs can then serve as
339 substrates for symbolic regression that extract interpretable laws from data [146, 147]. Conversely,
340 symbolic priors such as conservation laws, symmetry constraints, or algebraic invariants can be
341 encoded into the learning objective or architecture of the generative model, guiding it toward plausible
342 dynamics. This synergy enables a bidirectional interface: using symbolic structures to constrain and
343 interpret neural models, and using generative models to ground and test symbolic hypotheses. Such
344 integration holds transformative potential for discovering interpretable models of non-equilibrium
345 phenomena in physics, biology, and beyond.

346 **Neural ODEs/SDEs integrated Fokker–Planck–based model for non-equilibrium systems.**

347 Combining Fokker–Planck–based generative modeling with Neural ODEs/SDEs yields a uni-
348 fied, physics-informed framework for learning and sampling complex, non-equilibrium distribu-
349 tions [20, 148]. In this approach, both drift and diffusion processes are parameterized by neural
350 networks, and sample trajectories evolve under a learned stochastic process whose probability density
351 obeys Fokker–Planck dynamics. The model treats continuous time evolution as a normalizing flow,
352 enabling exact density tracking and invertible transformations. Training proceeds by maximizing the
353 data likelihood or matching score fields via the adjoint method so that the learned dynamics trans-
354 form a simple prior distribution into the target non-equilibrium steady state. This scheme captures
355 both microscopic particle evolution and macroscopic density evolution, ensuring generated samples
356 respect physical conservation laws and transient behaviors. Applications span plasma fluctuation
357 synthesis [149] and simulating many-body dynamics [150]. By fusing Fokker–Planck dynamics
358 with continuous-time neural solvers, this hybrid strategy offers a principled path for high-fidelity,
359 data-driven simulation and control of driven, dissipative systems.

Introducing non-equilibrium physical quantities and laws into generative models. Though diffusion models can be described without non-equilibrium dynamics, non-equilibrium thermodynamics yields crucial insights. Physical quantities, such as free energy, score functions, mutual information [151], and entropy production rate, map exactly onto diffusion metrics, enabling effective solutions for real-world tasks such as high-dimensional density estimation [152]. These mappings facilitate tackling non-equilibrium challenges by calibrating diffusion parameters and improving sampling efficiency. Moreover, established non-equilibrium theories guide the design of noise schedules and denoising functions: for example, Ikeda et al. [153] quantified a trade-off between entropy production and generation error—measured via Wasserstein distance—informing the development of efficient diffusion mechanisms. Moreover, it is potentially viable to optimize generative sampling by incorporating dissipated work and non-equilibrium work theorems into model training. Batch-estimate average dissipated work or Jarzynski objectives [54], dynamically adjust sampling drift during training, and apply Crooks path-probability regularization [56, 154] to accelerate convergence and improve efficiency.

Non-equilibrium physics for memory augmentation. Memory augmentation in generative models leverages non-equilibrium physics by embedding history-dependent feedback into latent dynamics. One approach uses delay kernels that convolve past latent states with learnable weights, capturing viscoelastic and hysteretic behavior. Alternatively, fractional derivative operators [155] introduce power-law memory across long timescales, modeling persistent correlations in fluctuation–dissipation processes. Autoregressive latent structures incorporate time lags, enabling the model to reference past states and approximate non-Markovian drift [156–158]. These mechanisms equip generative models with long-range memory, enabling accurate simulation of slow relaxation and driven steady states while preserving invertibility.

5 Alternative Views

While we advocate for integrating non-equilibrium statistical mechanics into generative modeling, several alternative perspectives merit discussion:

Empirical success of non-physics-inspired and equilibrium-based models might suggest that they are sufficient for scientific applications. This raises the question: why introduce non-equilibrium physics at all? We acknowledge these models’ success, but emphasize that the target system resides near a stationary regime or violates physical constraints. For dynamics systems, non-equilibrium models offer a principled route to temporal evolution; hence, it is more faithful to the temporal structure and entropy-producing nature of dynamic systems.

Computational complexity concerns. Non-equilibrium-inspired models could lead to significant training and sampling costs. Critics may view this as impractical for large-scale deployment. However, recent progress in neural solvers, adaptive sampling, and variational approximations has significantly reduced such burdens. Furthermore, the gains in modeling fidelity, especially for systems far from equilibrium, justify the additional complexity in many scientific contexts.

Questioning physical interpretability. Some may argue that non-equilibrium-based AI methods do not guarantee interpretability in learned representations. While interpretability does not automatically emerge, it would be beneficial from the fertile ground of non-equilibrium knowledge prior. In this way, non-equilibrium models provide scaffolds for extracting mechanistic insight, not just fitting data.

In summary, while equilibrium models and neural architectures without physical grounding have proven successful, they often fall short in representing the time-dependently non-equilibrium behaviors observed in real-world dynamical systems. Non-equilibrium generative models offer a complementary and increasingly necessary paradigm.

References

- [1] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.
- [2] Ludwig Boltzmann. *Weitere studien über das wärmeleichgewicht unter gasmolekülen*, volume 66. Aus der kk Hot-und Staatsdruckerei, 1872.
- [3] Josiah Willard Gibbs. *Elementary principles in statistical mechanics: developed with especial reference to the rational foundations of thermodynamics*. C. Scribner’s sons, 1902.
- [4] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.
- [5] Geoffrey E Hinton, Terrence J Sejnowski, and David H Ackley. *Boltzmann machines: Constraint satisfaction networks that learn*. Carnegie-Mellon University, Department of Computer Science Pittsburgh, PA, 1984.
- [6] David H Ackley, Geoffrey E Hinton, and Terrence J Sejnowski. A learning algorithm for boltzmann machines. *Cognitive science*, 9(1):147–169, 1985.
- [7] Geoffrey E Hinton, Terrence J Sejnowski, et al. Learning and relearning in boltzmann machines. *Parallel distributed processing: Explorations in the microstructure of cognition*, 1(282-317):2, 1986.
- [8] Geoffrey E Hinton. Training products of experts by minimizing contrastive divergence. *Neural computation*, 14(8):1771–1800, 2002.
- [9] Yann LeCun, Sumit Chopra, Raia Hadsell, Marc’Aurelio Ranzato, and Fu Jie Huang. A tutorial on energy-based learning. *Predicting structured data*, 1(0), 2006.
- [10] Yilun Du and Igor Mordatch. Implicit generation and modeling with energy based models. *Advances in Neural Information Processing Systems*, 32, 2019.
- [11] Will Grathwohl, Kuan-Chieh Wang, Jörn-Henrik Jacobsen, David Duvenaud, Mohammad Norouzi, and Kevin Swersky. Your classifier is secretly an energy based model and you should treat it like one. *arXiv preprint arXiv:1912.03263*, 2019.
- [12] Richard P Feynman. *Statistical mechanics: a set of lectures*. CRC press, 2018.
- [13] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [14] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *International conference on machine learning*, pages 1278–1286, 2014.
- [15] Durk P Kingma, Tim Salimans, Rafal Jozefowicz, Xi Chen, Ilya Sutskever, and Max Welling. Improved variational inference with inverse autoregressive flow. *Advances in neural information processing systems*, 29, 2016.
- [16] Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. *Advances in neural information processing systems*, 34:21696–21707, 2021.
- [17] Hadi Vafaii, Dekel Galor, and Jacob Yates. Poisson variational autoencoder. *Advances in Neural Information Processing Systems*, 37:44871–44906, 2024.
- [18] Erwin Schrödinger. Sur la théorie relativiste de l’électron et l’interprétation de la mécanique quantique. In *Annales de l’institut Henri Poincaré*, volume 2, pages 269–310, 1932.
- [19] Lars Onsager. Reciprocal relations in irreversible processes. ii. *Physical review*, 38(12):2265, 1931.
- [20] Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schrödinger bridge with applications to score-based generative modeling. *Advances in Neural Information Processing Systems*, 34:17695–17709, 2021.

- 451 [21] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsuper-
452 vised learning using nonequilibrium thermodynamics. In *International conference on machine*
453 *learning*, pages 2256–2265. PMLR, 2015.
- 454 [22] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances*
455 *in neural information processing systems*, 33:6840–6851, 2020.
- 456 [23] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and
457 Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv*
458 *preprint arXiv:2011.13456*, 2020.
- 459 [24] Giorgio Parisi. Correlation functions and computer simulations. *Nuclear Physics B*, 180(3):
460 378–384, 1981.
- 461 [25] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data
462 distribution. *Advances in neural information processing systems*, 32, 2019.
- 463 [26] Chieh-Hsin Lai, Yuhta Takida, Naoki Murata, Toshimitsu Uesaka, Yuki Mitsufuji, and Stefano
464 Ermon. Fp-diffusion: Improving score-based diffusion models by enforcing the underlying
465 score fokker-planck equation. In *International Conference on Machine Learning*, pages
466 18365–18398. PMLR, 2023.
- 467 [27] Yilun Xu, Ziming Liu, Max Tegmark, and Tommi Jaakkola. Poisson flow generative models.
468 *Advances in Neural Information Processing Systems*, 35:16782–16795, 2022.
- 469 [28] Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion
470 schrödinger bridge matching. *Advances in Neural Information Processing Systems*, 36, 2024.
- 471 [29] Francisco Vargas, Shreyas Padhy, Denis Blessing, and Nikolas Nüsken. Transport meets
472 variational inference: Controlled monte carlo diffusions. *arXiv preprint arXiv:2307.01050*,
473 2023.
- 474 [30] Suriyanarayanan Vaikuntanathan and Christopher Jarzynski. Escorted free energy simulations:
475 Improving convergence by reducing dissipation. *Physical Review Letters*, 100(19):190601,
476 2008.
- 477 [31] C. Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*,
478 78(14):2690–2693, 1997.
- 479 [32] Udo Seifert. Entropy production along a stochastic trajectory and an integral fluctuation
480 theorem. *Physical review letters*, 95(4):040602, 2005.
- 481 [33] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. In
482 *ICLR*, 2017.
- 483 [34] Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions.
484 In *NeurIPS*, 2018.
- 485 [35] Will Grathwohl, Ricky TQ Chen, Jesse Bettencourt, Ilya Sutskever, and David Duvenaud.
486 Ffjord: Free-form continuous dynamics for scalable reversible generative models. In *ICLR*,
487 2019.
- 488 [36] Aaron van den Oord, Nal Kalchbrenner, and Koray Kavukcuoglu. Pixel recurrent neural
489 networks. In *ICML*, 2016.
- 490 [37] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez,
491 Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information*
492 *processing systems*, 30, 2017.
- 493 [38] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. Improving language
494 understanding by generative pre-training. 2018.
- 495 [39] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil
496 Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in*
497 *neural information processing systems*, pages 2672–2680, 2014.

- 498 [40] Lev Davidovich Landau and Evgenii Mikhailovich Lifshitz. *Statistical Physics: Volume 5*,
499 volume 5. Elsevier, 2013.
- 500 [41] Ruslan Salakhutdinov, Andriy Mnih, and Geoffrey Hinton. Restricted boltzmann machines
501 for collaborative filtering. In *Proceedings of the 24th international conference on Machine*
502 *learning*, pages 791–798, 2007.
- 503 [42] Giovanni Gallavotti. *Statistical mechanics: A short treatise*. Springer Science & Business
504 Media, 1999.
- 505 [43] Ruiqi Gao, Erik Nijkamp, Diederik P Kingma, Zhen Xu, Andrew M Dai, and Ying Nian
506 Wu. Flow contrastive estimation of energy-based models. In *Proceedings of the IEEE/CVF*
507 *Conference on Computer Vision and Pattern Recognition*, pages 7518–7528, 2020.
- 508 [44] Erik Nijkamp, Mitch Hill, Tian Han, Song-Chun Zhu, and Ying Nian Wu. On the anatomy of
509 mcmc-based maximum likelihood learning of energy-based models. In *Proceedings of the*
510 *AAAI Conference on Artificial Intelligence*, volume 34, pages 5272–5280, 2020.
- 511 [45] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and
512 Edward Teller. Equation of state calculations by fast computing machines. *The journal of*
513 *chemical physics*, 21(6):1087–1092, 1953.
- 514 [46] W. Keith Hastings. Monte carlo sampling methods using markov chains and their applications.
515 *Biometrika*, 57(1):97–109, 1970.
- 516 [47] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian
517 restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*,
518 PAMI-6(6):721–741, 1984.
- 519 [48] Aaron Lou and Stefano Ermon. Reflected diffusion models. In *International Conference on*
520 *Machine Learning*, pages 22675–22701. PMLR, 2023.
- 521 [49] Jonathan Ho, Xi Chen, Aravind Srinivas, Yan Duan, and Pieter Abbeel. Flow++: Improving
522 flow-based generative models with variational dequantization and architecture design. In
523 *International conference on machine learning*, pages 2722–2730. PMLR, 2019.
- 524 [50] Qinsheng Zhang and Yongxin Chen. Diffusion normalizing flow. *Advances in neural informa-*
525 *tion processing systems*, 34:16280–16291, 2021.
- 526 [51] Nikita Gushchin, Sergei Kholkin, Evgeny Burnaev, and Alexander Korotin. Light and optimal
527 schrödinger bridge matching. In *Forty-first International Conference on Machine Learning*,
528 2024.
- 529 [52] Ilya Prigogine. Time, structure, and fluctuations. *Science*, 201(4358):777–785, 1978.
- 530 [53] Ilya Prigogine. *Non-equilibrium statistical mechanics*. Courier Dover Publications, 2017.
- 531 [54] Christopher Jarzynski. Equilibrium free-energy differences from nonequilibrium measure-
532 ments: A master-equation approach. *Physical Review E*, 56(5):5018, 1997.
- 533 [55] Radford M Neal. Annealed importance sampling. *Statistics and computing*, 11:125–139, 2001.
- 534 [56] Gavin E Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation
535 for free energy differences. *Physical Review E*, 60(3):2721, 1999.
- 536 [57] Ulf Grenander and Michael I Miller. Representations of knowledge in complex systems.
537 *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(4):549–581, 1994.
- 538 [58] Bernt Øksendal. *Stochastic differential equations*. Springer, 2003.
- 539 [59] Teo Deveney, Jan Stanczuk, Lisa Maria Kreusser, Chris Budd, and Carola-Bibiane Schönlieb.
540 Closing the ode-sde gap in score-based diffusion models through the fokker-planck equation.
541 *arXiv preprint arXiv:2311.15996*, 2023.

- 542 [60] Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural*
543 *computation*, 23(7):1661–1674, 2011.
- 544 [61] Yang Song, Sahaj Garg, Jiaxin Shi, and Stefano Ermon. Sliced score matching: A scalable
545 approach to density and score estimation. In *Uncertainty in Artificial Intelligence*, pages
546 574–584. PMLR, 2020.
- 547 [62] Ziming Liu, Di Luo, Yilun Xu, Tommi Jaakkola, and Max Tegmark. Genphys: From physical
548 processes to generative models. *arXiv preprint arXiv:2304.02637*, 2023.
- 549 [63] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow
550 matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.
- 551 [64] Christian Léonard. A survey of the schrödinger problem and some of its connections with
552 optimal transport. *arXiv preprint arXiv:1308.0215*, 2013.
- 553 [65] Yongxin Chen, Tryphon T Georgiou, and Michele Pavon. Optimal transport in systems and
554 control. *Annual Review of Control, Robotics, and Autonomous Systems*, 4(1):89–113, 2021.
- 555 [66] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick,
556 Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a
557 constrained variational framework. In *International conference on learning representations*,
558 2017.
- 559 [67] Michael I Jordan, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul. An
560 introduction to variational methods for graphical models. *Machine learning*, 37(2):183–233,
561 1999.
- 562 [68] David JC MacKay. *Information theory, inference and learning algorithms*. Cambridge
563 university press, 2003.
- 564 [69] Aaron van den Oord, Sander Dieleman, Heiga Zen, et al. Wavenet: A generative model for
565 raw audio. *arXiv preprint arXiv:1609.03499*, 2016.
- 566 [70] Mark Chen, Alec Radford, Rewon Child, Jeffrey Wu, Heewoo Jun, David Luan, and Ilya
567 Sutskever. Generative pretraining from pixels. In *International conference on machine*
568 *learning*, pages 1691–1703. PMLR, 2020.
- 569 [71] Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Jauvin. A neural probabilistic
570 language model. *Journal of Machine Learning Research*, 3:1137–1155, 2003.
- 571 [72] Ilya Sutskever, James Martens, and Geoffrey E Hinton. Generating text with recurrent neural
572 networks. In *Proceedings of the 28th international conference on machine learning (ICML-11)*,
573 pages 1017–1024, 2011.
- 574 [73] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhari-
575 wal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language
576 models are few-shot learners. *Advances in neural information processing systems*, 33:1877–
577 1901, 2020.
- 578 [74] Lijun Yu, Yong Cheng, Kihyuk Sohn, José Lezama, Han Zhang, Huiwen Chang, Alexander G
579 Hauptmann, Ming-Hsuan Yang, Yuan Hao, Irfan Essa, et al. Magvit: Masked generative video
580 transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern*
581 *Recognition*, pages 10459–10469, 2023.
- 582 [75] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with
583 deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
- 584 [76] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity
585 natural image synthesis. *arXiv preprint arXiv:1809.11096*, 2018.
- 586 [77] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative
587 adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and*
588 *pattern recognition*, pages 4401–4410, 2019.

- [78] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020.
- [79] Ruikun Li, Huandong Wang, Qingmin Liao, and Yong Li. Predicting the energy landscape of stochastic dynamical system via physics-informed self-supervised learning. *arXiv preprint arXiv:2502.16828*, 2025.
- [80] Marco Federici, Patrick Forré, Ryota Tomioka, and Bastiaan S Veeling. Latent representation and simulation of markov processes via time-lagged information bottleneck. *arXiv preprint arXiv:2309.07200*, 2023.
- [81] Jiazhen Liu, Shengda Huang, Nathaniel M Aden, Neil F Johnson, and Chaoming Song. Emergence of polarization in coevolving networks. *Physical Review Letters*, 130(3):037401, 2023.
- [82] Jiazhen Liu, Kunal Tamang, Dashun Wang, and Chaoming Song. Correlated growth of causal networks. *arXiv preprint arXiv:2412.16647*, 2024.
- [83] Jiazhen Liu, Nathaniel M Aden, Debasish Sarker, and Chaoming Song. Dynamical phase transitions in non-equilibrium networks. *arXiv preprint arXiv:2412.06704*, 2024.
- [84] Tom Chou, Kirone Mallick, and Royce KP Zia. Non-equilibrium statistical mechanics: from a paradigmatic model to biological transport. *Reports on Progress in Physics*, 74(11):116601, 2011. doi: 10.1088/0034-4885/74/11/116601.
- [85] Peter Reimann. Brownian motors: noisy transport far from equilibrium. *Physics Reports*, 361(2-4):57–265, 2002. doi: 10.1016/S0370-1573(01)00081-3.
- [86] Ying Tang, Jing Liu, Jiang Zhang, and Pan Zhang. Learning nonequilibrium statistical mechanics and dynamical phase transitions. *Nature Communications*, 15(1):1117, 2024.
- [87] Huili Liao, Xiaoyu Huang, Siyu Xiang, et al. Ai-nerd: Elucidation of relaxation dynamics beyond equilibrium via deep learning. *Nature Communications*, 15(1):4938, 2024. doi: 10.1038/s41467-024-45667-7.
- [88] Lars Holdijk, Yuanqi Du, Ferry Hooft, Priyank Jaini, Berend Ensing, and Max Welling. Stochastic optimal control for collective variable free sampling of molecular transition paths. *Advances in Neural Information Processing Systems*, 36:79540–79556, 2023.
- [89] Benjamin Kurt Miller, Ricky TQ Chen, Anuroop Sriram, and Brandon M Wood. Flowmm: Generating materials with riemannian flow matching. In *Forty-first International Conference on Machine Learning*, 2024.
- [90] Tian Xie and Jeffrey C Grossman. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical review letters*, 120(14):145301, 2018.
- [91] Rui Jiao, Wenbing Huang, Peijia Lin, Jiaqi Han, Pin Chen, Yutong Lu, and Yang Liu. Crystal structure prediction by joint equivariant diffusion. *Advances in Neural Information Processing Systems*, 36:17464–17497, 2023.
- [92] Nicola De Cao and Thomas Kipf. Molgan: An implicit generative model for small molecular graphs. *arXiv preprint arXiv:1805.11973*, 2018.
- [93] Haitao Lin, Yufei Huang, Odin Zhang, Siqi Ma, Meng Liu, Xuanjing Li, Lirong Wu, Jishui Wang, Tingjun Hou, and Stan Z Li. Diffbp: Generative diffusion of 3d molecules for target protein binding. *Chemical Science*, 16(3):1417–1431, 2025.
- [94] Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. Geodiff: A geometric diffusion model for molecular conformation generation. *arXiv preprint arXiv:2203.02923*, 2022.

- [95] Claudio Zeni, Robert Pinsler, Daniel Zügner, Andrew Fowler, Matthew Horton, Xiang Fu, Zilong Wang, Aliaksandra Shysheya, Jonathan Crabbé, Shoko Ueda, et al. A generative model for inorganic materials design. *Nature*, pages 1–3, 2025.
- [96] Daniel Levy, Siba Smarak Panigrahi, Sékou-Oumar Kaba, Qiang Zhu, Kin Long Kelvin Lee, Mikhail Galkin, Santiago Miret, and Siamak Ravanbakhsh. Symmcd: Symmetry-preserving crystal generation with diffusion models. *arXiv preprint arXiv:2502.03638*, 2025.
- [97] William Gilpin. Model scale versus domain knowledge in statistical forecasting of chaotic systems. *Physical Review Research*, 5(4):043252, 2023.
- [98] Ruikun Li, Huandong Wang, Jinghua Piao, Qingmin Liao, and Yong Li. Predicting long-term dynamics of complex networks via identifying skeleton in hyperbolic space. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1655–1666, 2024.
- [99] Yang Ye, Abhishek Pandey, Carolyn Bawden, Dewan Md Sumsuzzman, Rimpi Rajput, Affan Shoukat, Burton H Singer, Seyed M Moghadas, and Alison P Galvani. Integrating artificial intelligence with mechanistic epidemiological modeling: a scoping review of opportunities and challenges. *Nature Communications*, 16(1):1–18, 2025.
- [100] Jingtao Ding, Chang Liu, Yu Zheng, Yunke Zhang, Zihan Yu, Ruikun Li, Hongyi Chen, Jinghua Piao, Huandong Wang, Jiazhen Liu, et al. Artificial intelligence for complex network: Potential, methodology and application. *arXiv preprint arXiv:2402.16887*, 2024.
- [101] Aditya Grover, Aaron Zweig, and Stefano Ermon. Graphite: Iterative generative modeling of graphs. In *International conference on machine learning*, pages 2434–2444. PMLR, 2019.
- [102] Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.
- [103] Tomislav Stankovski, Tiago Pereira, Peter VE McClintock, and Aneta Stefanovska. Coupling functions: Universal insights into dynamical interaction mechanisms. *Reviews of Modern Physics*, 89(4):045001, 2017.
- [104] Tian-Tian Gao and Gang Yan. Autonomous inference of complex network dynamics from incomplete and noisy data. *Nature Computational Science*, 2:160–168, 2022.
- [105] Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge university press, 2015.
- [106] Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, Cambridge, 2000.
- [107] Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974. doi: 10.1037/h0037350.
- [108] Jakob Zscheischler, Sonia I. Seneviratne, Thomas R. Karl, and et al. Understanding and managing connected extreme events. *Nature Climate Change*, 10:611–621, 2020. doi: 10.1038/s41558-020-0790-4.
- [109] Satyaki Kundu, Tapas Bar, Rajesh Kumble Nayak, and Bhavtosh Bansal. Critical slowing down at the abrupt mott transition: When the first-order phase transition becomes zeroth-order and looks like second-order. *Physical Review Letters*, 124(9):095703, 2020. doi: 10.1103/PhysRevLett.124.095703.
- [110] Jan Korbelt, Shlomo Havlin, and Stefan Thurner. Microscopic origin of abrupt mixed-order phase transitions. *Nature Communications*, 16(1):2628, 2025. doi: 10.1038/s41467-025-57007-1.
- [111] Stefano Boccaletti, Juan A. Almendral, Shuguang Guan, Iacopo Leyva, Zonghua Liu, Iker Sendiña-Nadal, Zhen Wang, and Yamir Moreno. Explosive transitions in complex networks’ structure and dynamics: percolation and synchronization. *Physics Reports*, 660:1–94, 2016. doi: 10.1016/j.physrep.2016.10.004.

- [112] Denis J Evans, Ezechiel Godert David Cohen, and Gary P Morriss. Probability of second law violations in shearing steady states. *Physical review letters*, 71(15):2401, 1993.
- [113] Giovanni Gallavotti and Ezechiel Godert David Cohen. Dynamical ensembles in nonequilibrium statistical mechanics. *Physical review letters*, 74(14):2694, 1995.
- [114] Magnus Petersen, Gemma Roig, and Roberto Covino. Dynamicsdiffusion: Generating and rare event sampling of molecular dynamic trajectories using diffusion models.
- [115] Dvir Samuel, Rami Ben-Ari, Simon Raviv, Nir Darshan, and Gal Chechik. Generating images of rare concepts using pre-trained diffusion models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 4695–4703, 2024.
- [116] Huandong Wang, Huan Yan, Can Rong, Yuan Yuan, Fenyu Jiang, Zhenyu Han, Hongjie Sui, Depeng Jin, and Yong Li. Multi-scale simulation of complex systems: a perspective of integrating knowledge and data. *ACM Computing Surveys*, 56(12):1–38, 2024.
- [117] Riccardo Togni, Andrea Cimarelli, and Elisabetta De Angelis. Physical and scale-by-scale analysis of rayleigh–bénard convection. *Journal of Fluid Mechanics*, 782:380–404, 2015.
- [118] Jinghai Li. *Particle-fluid two-phase flow: the energy-minimization multi-scale method*. Metallurgical Industry Press, 1994.
- [119] Wei Ge and Jinghai Li. Physical mapping of fluidization regimes—the emms approach. *Chemical Engineering Science*, 57(18):3993–4004, 2002.
- [120] Yang Hsia, Rubul Mout, William Sheffler, Natasha I Edman, Ivan Vulovic, Young-Jun Park, Rachel L Redler, Matthew J Bick, Asim K Bera, Alexis Courbet, et al. Design of multi-scale protein complexes by hierarchical building block fusion. *Nature Communications*, 12(1):2294, 2021.
- [121] Xinyao Fan, Yueying Wu, Chang Xu, Yuhao Huang, Weiqing Liu, and Jiang Bian. Mg-tds: Multi-granularity time series diffusion models with guided learning process. In *The Twelfth International Conference on Learning Representations*, 2024.
- [122] Lizao Li, Robert Carver, Ignacio Lopez-Gomez, Fei Sha, and John Anderson. Generative emulation of weather forecast ensembles with diffusion models. *Science Advances*, 10(13): eadk4489, 2024.
- [123] Pan Du, Meet Hemant Parikh, Xiantao Fan, Xin-Yang Liu, and Jian-Xun Wang. Conditional neural field latent diffusion model for generating spatiotemporal turbulence. *Nature Communications*, 15(1):10416, 2024.
- [124] Kanta Masuki and Yuto Ashida. Generative diffusion model with inverse renormalization group flows. *arXiv preprint arXiv:2501.09064*, 2025.
- [125] Yuchen Zhu, Tianrong Chen, Evangelos A Theodorou, Xie Chen, and Molei Tao. Quantum state generation with structure-preserving diffusion model. *arXiv preprint arXiv:2404.06336*, 2024.
- [126] Zijie Huang, Yizhou Sun, and Wei Wang. Coupled graph ode for learning interacting system dynamics. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, pages 705–715, 2021.
- [127] Kevin E Wu, Kevin K Yang, Rianne van den Berg, Sarah Alamdari, James Y Zou, Alex X Lu, and Ava P Amini. Protein structure generation via folding diffusion. *Nature communications*, 15(1):1059, 2024.
- [128] Vignesh Ram Somnath, Charlotte Bunne, and Andreas Krause. Multi-scale representation learning on proteins. *Advances in Neural Information Processing Systems*, 34:25244–25255, 2021.
- [129] James M Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7(9–12):1393–1512, 1986. doi: 10.1016/0270-0255(86)90088-6.

- [130] Bernhard Schölkopf et al. Causal inference using graphical models with the r package pcalg. In *Proceedings of the 29th International Conference on Machine Learning (ICML)*, 2012.
- [131] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of Causal Inference: Foundations and Learning Algorithms*. MIT Press, 2017.
- [132] Yoshua Bengio, Tristan Deleu, Nasim Rahaman, Nan Ke, Sebastien Lachapelle, Olexa Bilaniuk, Anirudh Goyal, and Chris Pal. A meta-transfer objective for learning to disentangle causal mechanisms. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, 2020.
- [133] Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds. *arXiv preprint arXiv:1802.08219*, 2018.
- [134] Samuel Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian neural networks. *Advances in neural information processing systems*, 32, 2019.
- [135] Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E Smidt, and Boris Kozinsky. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature communications*, 13(1): 2453, 2022.
- [136] Hidenori Tanaka and Daniel Kunin. Noether’s learning dynamics: Role of symmetry breaking in neural networks. *Advances in Neural Information Processing Systems*, 34:25646–25660, 2021.
- [137] Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pages 2990–2999. PMLR, 2016.
- [138] Risi Kondor and Shubhendu Trivedi. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In *International conference on machine learning*, pages 2747–2755. PMLR, 2018.
- [139] Marc Finzi, Samuel Stanton, Pavel Izmailov, and Andrew Gordon Wilson. Generalizing convolutional neural networks for equivariance to lie groups on arbitrary continuous data. In *International Conference on Machine Learning*, pages 3165–3176. PMLR, 2020.
- [140] Jaideep Pathak, Brian Hunt, Michelle Girvan, Zhixin Lu, and Edward Ott. Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. *Physical review letters*, 120(2):024102, 2018.
- [141] Frank Noé, Simon Olsson, Jonas Köhler, and Hao Wu. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457):eaaw1147, 2019.
- [142] Charlotte Bunne, Ya-Ping Hsieh, Marco Cuturi, and Andreas Krause. The schrödinger bridge between gaussian measures has a closed form. In *Proceedings of the 40th International Conference on Machine Learning*, pages 1234–1243, 2023.
- [143] Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.
- [144] Ruikun Li, Jingwen Cheng, Huandong Wang, Qingmin Liao, and Yong Li. Predicting the dynamics of complex system via multiscale diffusion autoencoder. *arXiv preprint arXiv:2505.02450*, 2025.
- [145] Nour Makke and Sanjay Chawla. Interpretable scientific discovery with symbolic regression: a review. *Artificial Intelligence Review*, 57(1):2, 2024.
- [146] Zihan Yu, Jingtao Ding, and Yong Li. Symbolic regression via mdlformer-guided search: from minimizing prediction error to minimizing description length. In *The Thirteenth International Conference on Learning Representations*, 2025.

- [147] Hongzhi Shi, Jingtao Ding, Yufan Cao, Li Liu, Yong Li, et al. Learning symbolic models for graph-structured physical mechanism. In *The Eleventh International Conference on Learning Representations*, 2023.
- [148] Macheng Shen and Chen Cheng. Neural sdes as a unified approach to continuous-domain sequence modeling. *arXiv preprint arXiv:2501.18871*, 2025.
- [149] Jan Trieschmann, Luca Vialetto, and Tobias Gergs. Machine learning for advancing low-temperature plasma modeling and simulation. *Journal of Micro/Nanopatterning, Materials, and Metrology*, 22(4):041504–041504, 2023.
- [150] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017.
- [151] Xianghao Kong, Rob Brekelmans, and Greg Ver Steeg. Information-theoretic diffusion. In *The Eleventh International Conference on Learning Representations*.
- [152] Nicholas M Boffi and Eric Vanden-Eijnden. Deep learning probability flows and entropy production rates in active matter. *Proceedings of the National Academy of Sciences*, 121(25):e2318106121, 2024.
- [153] Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2009.
- [154] Rep Kubo. The fluctuation-dissipation theorem. *Reports on progress in physics*, 29(1):255, 1966.
- [155] Wenjun Cui, Qiyu Kang, Xuhao Li, Kai Zhao, Wee Peng Tay, Weihua Deng, and Yidong Li. Neural variable-order fractional differential equation networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 16109–16117, 2025.
- [156] IA Luchnikov, SV Vintskevich, DA Grigoriev, and SN Filippov. Machine learning non-markovian quantum dynamics. *Physical review letters*, 124(14):140502, 2020.
- [157] Qingchao Jiang, Xiaoming Fu, Shifu Yan, Runlai Li, Wenli Du, Zhixing Cao, Feng Qian, and Ramon Grima. Neural network aided approximation and parameter inference of non-markovian models of gene expression. *Nature communications*, 12(1):2618, 2021.
- [158] Ruikun Li, Huandong Wang, and Yong Li. Learning slow and fast system dynamics via automatic separation of time scales. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4380–4390, 2023.
- [159] Chence Shi, Shitong Luo, Minkai Xu, and Jian Tang. Learning gradient fields for molecular conformation generation. In *ICML*, pages 9558–9568. PMLR, 2021.
- [160] Shitong Luo, Chence Shi, Minkai Xu, and Jian Tang. Predicting molecular conformation via dynamic graph score matching. *NeurIPS*, 34:19784–19795, 2021.
- [161] Bowen Jing, Gabriele Corso, Jeffrey Chang, Regina Barzilay, and Tommi Jaakkola. Torsional diffusion for molecular conformer generation. *Advances in Neural Information Processing Systems*, 35:24240–24253, 2022.
- [162] Namrata Anand and Tudor Achim. Protein structure and sequence generation with equivariant denoising diffusion probabilistic models. *arXiv preprint arXiv:2205.15019*, 2022.
- [163] Shitong Luo, Yufeng Su, Xingang Peng, Sheng Wang, Jian Peng, and Jianzhu Ma. Antigen-specific antibody design and optimization with diffusion-based generative models for protein structures. *Advances in Neural Information Processing Systems*, 35:9754–9767, 2022.
- [164] Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976):1089–1100, 2023.

- [165] John B Ingraham, Max Baranov, Zak Costello, Karl W Barber, Wujie Wang, Ahmed Ismail, Vincent Frappier, Dana M Lord, Christopher Ng-Thow-Hing, Erik R Van Vlack, et al. Illuminating protein space with a programmable generative model. *Nature*, 623(7989):1070–1078, 2023.
- [166] Jin Sub Lee, Jisun Kim, and Philip M Kim. Score-based generative modeling for de novo protein design. *Nature Computational Science*, 3(5):382–392, 2023.
- [167] Hyungjin Chung, Eun Sun Lee, and Jong Chul Ye. Mr image denoising and super-resolution using regularized reverse diffusion. *IEEE Transactions on Medical Imaging*, 42(4):922–934, 2022.
- [168] Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. *ICLR*, 2022.
- [169] Ajil Jalal, Marius Arvinte, Giannis Daras, Eric Price, Alexandros G Dimakis, and Jon Tamir. Robust compressed sensing mri with deep generative priors. *NeurIPS*, 34:14938–14954, 2021.
- [170] Hyungjin Chung and Jong Chul Ye. Score-based diffusion models for accelerated mri. *Medical image analysis*, 80:102479, 2022.
- [171] Chentao Cao, Zhuo-Xu Cui, Yue Wang, Shaonan Liu, Taijin Chen, Hairong Zheng, Dong Liang, and Yanjie Zhu. High-frequency space diffusion model for accelerated mri. *IEEE Transactions on Medical Imaging*, 2024.
- [172] Cheng Peng, Pengfei Guo, S Kevin Zhou, Vishal M Patel, and Rama Chellappa. Towards performant and reliable undersampled mr reconstruction via diffusion model sampling. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 623–633. Springer, 2022.
- [173] Yutong Xie and Quanzheng Li. Measurement-conditioned denoising diffusion probabilistic model for under-sampled medical image reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 655–664. Springer, 2022.
- [174] Dule Shu, Zijie Li, and Amir Barati Farimani. A physics-informed diffusion model for high-fidelity flow field reconstruction. *Journal of Computational Physics*, 478:111972, 2023.
- [175] Jiajun Hu, Zhen Lu, and Yue Yang. Generative prediction of flow field based on the diffusion model. *arXiv preprint arXiv:2407.00735*, 2024.
- [176] Ilan Price, Alvaro Sanchez-Gonzalez, Ferran Alet, Tom R Andersson, Andrew El-Kadi, Dominic Masters, Timo Ewalds, Jacklynn Stott, Shakir Mohamed, Peter Battaglia, et al. Probabilistic weather forecasting with machine learning. *Nature*, 637(8044):84–90, 2025.
- [177] Mariia Drozdova, Vitaliy Kinakh, Omkar Bait, Olga Taran, Erica Lastufka, Miroslava Dessauges-Zavadsky, Taras Holotyak, Daniel Schaerer, and Slava Voloshynovskiy. Radio-astronomical image reconstruction with a conditional denoising diffusion model. *Astronomy & Astrophysics*, 683:A105, 2024.
- [178] Renato Sortino, Thomas Cecconello, Andrea DeMarco, Giuseppe Fiameni, Andrea Pilzer, Daniel Magro, Andrew M Hopkins, Simone Riggi, Eva Sciacca, Adriano Ingallinera, et al. Rad-iff: controllable diffusion models for radio astronomical maps generation. *IEEE Transactions on Artificial Intelligence*, 2024.

A Representative cases of generative models driven by non-equilibrium stochastic dynamics

Random variables $\mathbf{x}_t$ that evolve according to a stochastic process can be modeled as an Ito process given by

$$d\mathbf{x}_t = \mu(\mathbf{x}_t, t)dt + \sigma(\mathbf{x}_t, t)dW_t, \quad (10)$$

where μ represents the drift term, characterizing the deterministic interactions or correlations within the system or data, and σ describes the diffusion term, capturing the inherent stochastic fluctuations. Here, W_t denotes a standard Wiener process. The corresponding partial differential equation, known as the Fokker–Planck equation, can be derived from Equation (10) to describe the evolution of the probability density associated with $\mathbf{x}_t$ (see Appendix A for the detailed derivation).

$$\frac{\partial P(\mathbf{x}, t)}{\partial(t)} = -\frac{\partial}{\partial \mathbf{x}}[\mu(\mathbf{x}, t)P(\mathbf{x}, t)] + \frac{\partial^2}{\partial \mathbf{x}^2}[D(\mathbf{x}, t)P(\mathbf{x}, t)], \quad (11)$$

where probability distribution $P(\mathbf{x}, t)$ is the target of the generative model, and the diffusion coefficient $D(\mathbf{x}, t) = \frac{\sigma(\mathbf{x}_t, t)^2}{2}$. It is noteworthy that the purpose of the generative model is to generate targeted distributions $P(\mathbf{x}, t)$ from real data samples $P_{real}(\mathbf{x})$, where $P_{real}(\mathbf{x}) = P(\mathbf{x}, t = 0) = P(\mathbf{x})\delta(t)$ corresponds to the initial condition of the stochastic process.

Example 1: Diffusion Generative Model.

For the learning process, Diffusion model chooses the noise levels $\{\sigma_t\}_{t=1}^T$ as the diffusion coefficients of the discrete Markov process. When $T \rightarrow \infty$, it converges to a pure continuous diffusion process where the diffusion coefficient depends on time t ,

$$d\mathbf{x}_t = \sigma(\mathbf{x}_t, t)dW_t. \quad (12)$$

The corresponding Fokker-Planck equation is

$$\frac{\partial P(\mathbf{x}, t)}{\partial(t)} = \frac{\partial^2}{\partial \mathbf{x}^2}[D(t)P(\mathbf{x}, t)], \quad (13)$$

where $D = \frac{\sigma(\mathbf{x}_t, t)^2}{2}$ and initial condition $P(\mathbf{x}, t = 0) = P(\mathbf{x})_{real}$. Diffusion generative model usually let the noise levels $\{\sigma_t\}_{t=1}^T$ as a geometric positive sequence, satisfying $\frac{\sigma_t}{\sigma_{t+1}} = C_0 > 1$, where C_0 is a constant. If this ratio $C_0 \sim 1$, then the prior distribution $P(\mathbf{x}, t \rightarrow \infty) = P(\mathbf{x}, T)$ can be analytically found to follow a Gaussian distribution asymptotically. For the more general cases, $P(\mathbf{x}, t)$ can be rigorously solved by Green’s function method with $G(\mathbf{x}, t|\mathbf{x}, t')$, ($t \geq t'$), corresponding to the transition kernel in the learning process. The nature of stochastic dynamics guarantee that $P(\mathbf{x}, T)$ is uncorrelated with $P(\mathbf{x}, t = 0) = P(\mathbf{x})_{real}$, which has been proved in previous work [62].

For the process of generating samples, it satisfies a reverse process of Equations (12) and (13)

$$d\mathbf{x}_t = -D(t)\frac{\partial \log P(\mathbf{x}, t)}{\partial \mathbf{x}}dt + \sigma(\mathbf{x}_t, t)d\tilde{W}_t. \quad (14)$$

Note that $s(\mathbf{x}) := \frac{\partial \log P(\mathbf{x})}{\partial \mathbf{x}}$ is the score function in [25, 23], and $\tilde{W}_t$ is the time-reversal Wiener Process. Based on Equation (14), we can obtain the generating samples from the distribution following the below Fokker-Planck equation

$$\frac{\partial P(\mathbf{x}, t)}{\partial(t)} = \frac{\partial}{\partial \mathbf{x}}[s(\mathbf{x})P(\mathbf{x}, t)] + \frac{\partial^2}{\partial \mathbf{x}^2}[P(\mathbf{x}, t)]. \quad (15)$$

Example 2: Poisson flow generative model (PFGM).

As demonstrated in [27], the Poisson Flow Generative Model (PFGM) represents a specific instance of the stochastic Poisson flow process. The gradient flow ODE of PFGM can be seen as a non-diffusive Ito process characterized by a Poisson-field-driven force. In the more general case, the corresponding SDE for the PFGM generation process describes the stochastic dynamics of particle flow under an electric field:

$$d\mathbf{x}_t = -\frac{\partial \tilde{\psi}(\mathbf{x})}{\partial \mathbf{x}}dt + \sigma(\mathbf{x}_t, t)d\tilde{W}_t, \quad (16)$$

900 and the corresponding Fokker-Planck equation

$$\frac{\partial P(\mathbf{x}, t)}{\partial(t)} = -\frac{\partial}{\partial \mathbf{x}}[E(\mathbf{x})P(\mathbf{x}, t)] + \frac{\partial^2}{\partial \mathbf{x}^2}[D(\mathbf{x}, t)P(\mathbf{x}, t)], \quad (17)$$

901 where $E(\mathbf{x}) = -\frac{\partial \tilde{\psi}(\mathbf{x})}{\partial \mathbf{x}}$ represents the electric field derived from the Poisson equation. By com-
 902 paring Equations (16)–(17) with earlier formulations such as (14)–(15), it becomes clear that the
 903 electric field $E(\mathbf{x})$ serves as the counterpart to the score function $s(\mathbf{x})$ in the earlier example. This
 904 highlights a profound physical interpretation of the score-based method, where the score function
 905 $s(\mathbf{x})$ corresponds mathematically to the gradient of the potential field—essentially, the "free energy."
 906 This equivalence confirms that the generative process is consistent with the well-established Action
 907 Principle in physics. For the specific case of PFGM, the diffusion term σ approaches zero, allowing it
 908 to utilize an ODE-based framework.

909 PFGM's training process involves perturbing real data in a manner that, while consistent with a
 910 stochastic dynamic process, is not strictly limited to a diffusion process as described by the Ito
 911 formulation in Equation (10). Unlike traditional approaches that rely on Gaussian noise, PFGM
 912 allows for noise distributions beyond Gaussian. Nevertheless, the noise scales chosen in PFGM
 913 remain closely aligned with those in *example 1*, which emphasizes the importance of points near
 914 the real data in the overall generative process. As a result, if PFGM were to sample noise from a
 915 Gaussian distribution, its forward process would recover the form given by Equation (12).

916 **Example 3: Diffusion Schrödinger Bridge Model (DSBM).**

917 In comparison to *examples 1 & 2*, DSBM provides a more general generative framework that aligns
 918 with Equations (10)–(11). DSBM is rooted in the classic non-equilibrium statistical physics problem
 919 known as the Schrödinger Bridge Problem. This foundational problem seeks to characterize the non-
 920 equilibrium stochastic transportation process that transitions from an initial distribution $P(x, t = 0)$
 921 to a target distribution $P(x, t = T)$ at time T . The underlying dynamics can be described by the
 922 following SDE:

$$d\mathbf{x}_t = F(\mathbf{x}_t, t)dt + \sigma(\mathbf{x}_t, t)dW_t, \quad (18)$$

923 where the drifting field $F(\mathbf{x}_t, t) = f(\mathbf{x}_t, t) + \sigma(\mathbf{x}_t, t)^2 \frac{\partial \psi(\mathbf{x})}{\partial \mathbf{x}}$. The corresponding Fokker-Planck
 924 equation satisfies

$$\frac{\partial P(\mathbf{x}, t)}{\partial(t)} = -\frac{\partial}{\partial \mathbf{x}}[F(\mathbf{x}, t)P(\mathbf{x}, t)] + \frac{\partial^2}{\partial \mathbf{x}^2}[D(\mathbf{x}, t)P(\mathbf{x}, t)]. \quad (19)$$

925 The time-reversal process of Equations (18) and (19) satisfies

$$d\mathbf{x}_t = \tilde{F}(\mathbf{x}_t, t)dt + \sigma(\mathbf{x}_t, t)d\tilde{W}_t, \quad (20)$$

926 and

$$\frac{\partial P(\mathbf{x}, t)}{\partial(t)} = -\frac{\partial}{\partial \mathbf{x}}[\tilde{F}(\mathbf{x}, t)P(\mathbf{x}, t)] + \frac{\partial^2}{\partial \mathbf{x}^2}[D(\mathbf{x}, t)P(\mathbf{x}, t)], \quad (21)$$

927 where $\tilde{F}(\mathbf{x}, t) = f(\mathbf{x}_t, t) - \sigma(\mathbf{x}_t, t)^2 \frac{\partial \tilde{\psi}(\mathbf{x})}{\partial \mathbf{x}}$. Moreover, given the ensemble distribution $P(x, t = T)$
 928 at time T and the initial condition $P(x, t = 0)$, Equations (18-21) should simultaneously satisfy
 929 $\psi(\mathbf{x}_t, t = 0)\tilde{\psi}(\mathbf{x}_t, t = 0) = P(\mathbf{x}_t, t = 0) = P_{real}(x)$ and $\psi(\mathbf{x}_t, t = T)\tilde{\psi}(\mathbf{x}_t, t = T) = P(\mathbf{x}_t, t =$
 930 $T)$. DSBM reframes the task of generating the ensemble distribution as one of finding an optimal
 931 energy field, represented by ψ and $\tilde{\psi}$. In machine learning applications, obtaining a closed-form
 932 analytical solution for ψ is infeasible. As a result, the problem of identifying ψ and $\tilde{\psi}$ is treated as
 933 a reparameterization challenge. Notably, both diffusion models and PFGM can be seen as specific
 934 instances of DSBM: the diffusion model corresponds to a constant potential $\psi = c$, while PFGM is
 935 characterized by $\tilde{\psi} = E(\mathbf{x})$ and a non-zero diffusion term $\sigma \sim 0$.

936 **B Generating samples by the Fokker-Planck equation.**

937 The Fokker-Planck equation (Equation (11)) describes the distribution sample $P(\mathbf{x}, t)$ generating
 938 process, which can be derived from the Eq. (10). Starting with the Ito's lemma with an arbitrary
 939 integrable function $f(x)$, we obtain

$$df(\mathbf{x}_t) = \left\{ f'(\mathbf{x}_t)\mu(\mathbf{x}_t, t) + \frac{1}{2}f''(\mathbf{x}_t)\sigma(\mathbf{x}_t, t)^2 \right\} dt + f'(\mathbf{x}_t)\sigma(\mathbf{x}_t, t)dW_t. \quad (22)$$

940 The above equation leads to

$$f(\mathbf{x}_t) = f(\mathbf{x}_t; T = 0) + \int_0^t \left\{ f'(\mathbf{x}_t) \mu(\mathbf{x}_t, t) + \frac{1}{2} f''(\mathbf{x}_t) \sigma(\mathbf{x}_t, t)^2 \right\} dt + \int_0^t f'(\mathbf{x}_t) \sigma(\mathbf{x}_t, t) dW_t. \quad (23)$$

941 Next we define the probability density function $P(\mathbf{x}_t|t)$. Incorporating $P(\mathbf{x}_t|t)$ with Eq. (23) leads
942 to

$$\begin{aligned} \int_a^b P(\mathbf{x}_t|t) f(\mathbf{x}_t) d\mathbf{x}_t &= \int_a^b d\mathbf{x}_t P(\mathbf{x}_t|t) \int_0^T \left\{ f'(\mathbf{x}_t) \mu(\mathbf{x}_t, t) + \frac{1}{2} f''(\mathbf{x}_t) \sigma(\mathbf{x}_t, t)^2 \right\} dt \\ &\quad + \int_a^b d\mathbf{x}_t P(\mathbf{x}_t|t) \int_0^T f'(\mathbf{x}_t) \sigma(\mathbf{x}_t, t) dW_t + f_0(\mathbf{x}_t). \end{aligned} \quad (24)$$

943 Given the expectation of the randomness term of Ito process equal to zero

944 $\int_a^b \int_0^T P(\mathbf{x}_t|t) f'(\mathbf{x}_t) \sigma(\mathbf{x}_t) dW_t d\mathbf{x}_t = 0$

$$\int_a^b P(\mathbf{x}_t|t) f(\mathbf{x}_t) d\mathbf{x}_t = \int_a^b d\mathbf{x}_t P(\mathbf{x}_t|t) \int_0^T \left\{ f'(\mathbf{x}_t) \mu(\mathbf{x}_t, t) + \frac{1}{2} f''(\mathbf{x}_t) \sigma(\mathbf{x}_t)^2 \right\} dt + f_0(\mathbf{x}_t) \quad (25)$$

945 Then, we take time derivative on the both side of Eq.(25)

$$\frac{\partial}{\partial t} \int_a^b P(\mathbf{x}_t|t) f(\mathbf{x}_t) d\mathbf{x}_t = \int_a^b d\mathbf{x}_t P(\mathbf{x}_t|t) \left[f'(\mathbf{x}_t) \mu(\mathbf{x}_t, t) + \frac{1}{2} f''(\mathbf{x}_t) \sigma(\mathbf{x}_t)^2 \right].$$

946 By using the partial integral method, we obtain

$$\begin{aligned} \frac{\partial}{\partial t} \int_a^b P(\mathbf{x}_t|t) f(\mathbf{x}_t) d\mathbf{x}_t &= P(\mathbf{x}_t|t) \mu(\mathbf{x}_t, t) \Big|_a^b - \int_a^b d\mathbf{x}_t \frac{\partial [P(\mathbf{x}_t|t) \mu(\mathbf{x}_t, t)]}{\partial \mathbf{x}_t} f(\mathbf{x}_t) \\ &\quad + \frac{1}{2} P(\mathbf{x}_t|t) f'(\mathbf{x}_t) \sigma(\mathbf{x}_t, t)^2 \Big|_a^b - \frac{1}{2} \int_a^b d\mathbf{x}_t \frac{\partial (P(\mathbf{x}_t|t) \sigma(\mathbf{x}_t)^2)}{\partial \mathbf{x}_t} f'(\mathbf{x}_t). \end{aligned}$$

947 Notice that $P(\mathbf{x}_t|t) \mu(\mathbf{x}_t, t) \Big|_a^b = 0$ and $\frac{1}{2} P(\mathbf{x}_t|t) f'(\mathbf{x}_t) \sigma(\mathbf{x}_t, t)^2 \Big|_a^b = 0$. Therefore, we finally obtain

$$\frac{\partial}{\partial t} \int_a^b P(\mathbf{x}_t|t) f(\mathbf{x}_t) d\mathbf{x}_t = - \int_a^b d\mathbf{x}_t \frac{\partial [P(\mathbf{x}_t|t) \mu(\mathbf{x}_t, t)]}{\partial \mathbf{x}_t} f(\mathbf{x}_t) + \int_a^b d\mathbf{x}_t \frac{\partial^2 (P(\mathbf{x}_t|t) \sigma(\mathbf{x}_t)^2)}{\partial \mathbf{x}_t^2} f(\mathbf{x}_t). \quad (26)$$

948 The above equation is the integral formulation of Fokker-Planck equation. For the three representative
949 examples, their Fokker-Planck equation exhibits different drifting and diffusion behaviors, we
summarize them in the following table

Table 2: Summary of the drifting and diffusion terms of generative models.

Models	Forward Drifting	Forward Diffusion	Backward Drifting	Backward Diffusion
DGM	0	$\sigma(\mathbf{x}_t, t)$	$-\frac{\sigma(\mathbf{x}_t, t)^2}{2} s(\mathbf{x})$	$\sigma(\mathbf{x}_t, t)$
PFGM	NA	NA	$E(\mathbf{x})$	~ 0
SBM	$\sim \sigma(\mathbf{x}_t, t)^2 \frac{\partial \psi(\mathbf{x})}{\partial \mathbf{x}}$	$\sigma(\mathbf{x}_t, t)$	$\sim -\sigma(\mathbf{x}_t, t)^2 \frac{\partial \psi(\mathbf{x})}{\partial \mathbf{x}}$	$\sigma(\mathbf{x}_t, t)$

950

C Asymptotical independency of the initial state for the Brownian motion

The Brownian motion is a typical stationary independent incremental stochastic process, where $\mathbf{x}_{t_0}, \mathbf{x}_{t_1} - \mathbf{x}_{t_0}, \dots, \mathbf{x}_{t_n} - \mathbf{x}_{t_{n-1}}$ are independent and identical distributed for $t \in [0, \infty)$. This indicates that the (drifting) Brownian motion is characterized by the Green's function $G(\mathbf{x}, t | \mathbf{x}_0, 0)$, which corresponds to the transition probability of generative models. Note that only non-drifting process satisfy $G(\mathbf{x}, t | \mathbf{x}_0, 0) = \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{(\mathbf{x}-\mathbf{x}_0)^2}{2\sigma^2 t}}$, which is the special case for drifting Brownian motion. For the other drifting process, $\frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{(\mathbf{x}-\mathbf{x}_0)^2}{2\sigma^2 t}}$ is the first order term of their Green's function. Therefore, we have

$$G(\mathbf{x}, t | \mathbf{x}_0, 0) \sim \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{(\mathbf{x}-\mathbf{x}_0)^2}{2\sigma^2 t}} \quad (27)$$

Given the Fokker-Planck equation, we can always write down its analytical solution in an integral form by using the Greens function, following

$$P(\mathbf{x}, t) = \int d\mathbf{x} P(\mathbf{x}_0, 0) G(\mathbf{x}, t | \mathbf{x}_0, 0). \quad (28)$$

Considering a large $t = t_\infty \rightarrow \infty$, (27) indicates that $G \sim \frac{1}{\sqrt{2\pi\sigma^2 t}}$, which means the Green's function are governs by large time t . This further indicates

$$P(\mathbf{x}, t \rightarrow \infty) = \int d\mathbf{x} P(\mathbf{x}_0, 0) e^{-\frac{(\mathbf{x}-\mathbf{x}_0)^2}{2\sigma^2 t_\infty}} = \int d\mathbf{x} P(\mathbf{x}'_0, 0) e^{-\frac{(\mathbf{x}-\mathbf{x}'_0)^2}{2\sigma^2 t_\infty}}, \quad (29)$$

where $P(\mathbf{x}_0, 0)$ and $P(\mathbf{x}'_0, 0)$ are different initial distributions. (29) means the stationary state $P(\mathbf{x}, t \rightarrow \infty)$ of drifting Brownian motion are independent of their initial states because the Greens function's asymptotic behavior for is dominated by large t . This aligns with the requirement of generative model: the prior distribution $P(\mathbf{x}, t \rightarrow \infty)$ is independent of the real data $P_{real}(\mathbf{x})$.

D Experiment on a time-varying potential system

To directly compare the differences in generative models guided by the principles of equilibrium and non-equilibrium processes for generative modeling of complex systems, we modify and conduct tests on a 2-dimensional dynamic system in relevant literature [79, 80].

This system is a Printz potential well within a 2-dimensional bounded region, governed by the equation

$$V(t, x, y) = \cos\left(s \arctan(y, x) - \frac{\pi}{2}t\right) + 10\left(\sqrt{x^2 + y^2} - \frac{1}{2}\right), \quad (30)$$

$$dX_t = -\nabla V(X_t, t)dt + \sqrt{2\beta^{-1}}dW_t, \quad (31)$$

where $s = 5$ defines the number of potential wells around the origin and $\beta = 10$ controls the noise intensity. The time-dependent term in the potential function leads to a time-varying energy landscape with a period of $\pi/2s$. This simulates the system’s response to periodic external disturbances, serving as a simplified representation of real-world scenarios. The particle’s motion follows overdamped Langevin dynamics, with its drag term described by the potential energy landscape. As the potential energy changes over time, the particle’s equilibrium state is continuously perturbed, causing the system to remain in non-equilibrium transitions between different states.

Our generative modeling objective is to learn the time-varying energy field $V(t, x, y)$ and subsequently generate reliable particle distributions $p(x, t)$. Inspired by the distinctions between equilibrium and non-equilibrium processes, we adopt two approaches for modeling the energy field:

- **Equilibrium method** assumes that the particle distribution at each time instance adheres to a Boltzmann distribution defined by the energy [141]. We construct an approximately static energy field by statistically analyzing sample frequencies. The static energy field at each time instance is then used to supervise the training of a parameterized time-varying neural field $V_\theta(t, x, y)$. This process distills the continuously evolving energy field into a parameterized model for continuous sampling during testing.
- **Non-equilibrium method** focuses on learning the particle distribution as a function of time, $p(x|t)$. We employ a conditional diffusion model [25], incorporating the dynamic time t as an additional conditional input to the parameterized score function $f_\theta(x_n, n)$. This enables the model to learn the system’s time-varying energy gradient directly from continuous dynamic trajectories.

We uniformly sample 2,000 initial positions within the region $[-1, 1]^2$ and simulate a total duration of $T = 20.0$ with a unit time step of $dt = 0.02$. Half of these 2,000 trajectories are used for training, and the other half for testing. We use Jensen-Shannon divergence (JSD) as the error metric for generated particle trajectories.

E Generative models in various physics systems

Molecular System. Molecules are commonly represented as graphs, with nodes representing atoms and edges representing chemical bonds. Conformation prediction is a task that determines the three-dimensional (3D) coordinates of all atoms in a given molecule, which offers a more intrinsic representation of molecules. Traditionally, this task relies on molecular dynamics simulations, which sequentially update the coordinates of atoms based on the forces acting on each atom. This process can be linked to non-equilibrium thermodynamics by considering the score function as pseudo-forces that guide atoms toward high-likelihood regions. Shi et al. [159] propose ConfGS that uses a graph isomorphism network to estimate the score function and achieves state-of-the-art (SOTA) performance. However, it does not account for long-range atomic interactions like van der Waals forces. To address this, Luo et al. [160] propose to use the graph constructed dynamically based on spatial proximity, which improves performance in predicting protein side chains and multi-molecular complexes. These methods add noise to distance matrices, which can violate the triangular inequality or lead to negative values. Therefore, Xu et al. [94] propose to add noise to atomic coordinates instead of distance matrices, significantly improving the success rate. Jing et al. [161] suggest applying the diffusion and reverse processes solely to the chemical bond torsions since it has the most significant impact on conformations, outperforming traditional cheminformatics methods for the first time.

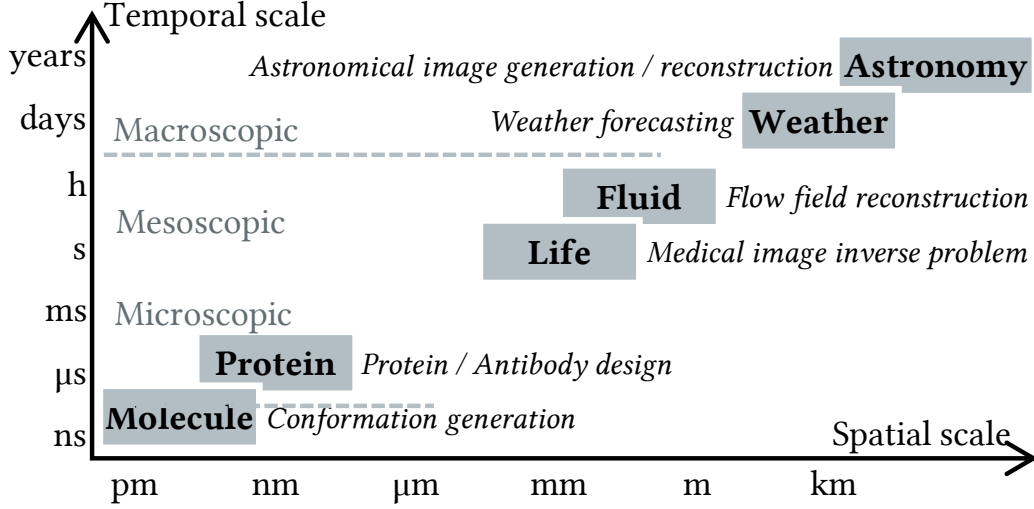


Figure 3: Physical systems invoke non-equilibrium thermodynamics and the associated problems.

Protein System. Proteins are biological macromolecules composed of amino acids. The *sequence* of amino acids determines its three-dimensional *structure*, which further dictates its *function*. The protein design task aims to generate protein sequences and/or structures with specific functions. Anand and Achim [162], Luo et al. [163], Watson et al. [164], Ingraham et al. [165] propose diffusion-based methods for generating protein sequences and structures, while Lee et al. [166] presents a score-based method. Anand and Achim [162] introduce a specialized diffusion scheme to generate amino acids sequences in discrete space and side chain orientations in non-Euclidean space. Luo et al. [163], Watson et al. [164] focus on generating proteins conditioned on richer prompts, such as specific antigen structures and functional descriptions. Meanwhile, Ingraham et al. [165] proposes an improved model architecture that reduces time complexity when generating proteins or protein complexes. Wu et al. [127] links the denoising diffusion process to the physics process by which proteins fold to minimize energy. Instead of working with atomic Cartesian coordinates, they perform the diffusion and reverse processes on bond angles and torsion between atoms in the protein backbone.

Life System. The observation of life systems heavily depends on medical imaging techniques such as magnetic resonance imaging (MRI) and computed tomography (CT), which capture data in Fourier space or sinogram space. However, obtaining full measurements often incurs significant time costs or excessive ionizing radiation exposure to patients. Therefore, it is essential to develop algorithms that can generate medical images from partial measurements, such as downsampled Fourier space data or sparse-view sinograms. Unlike deterministic supervised learning methods, diffusion model-based generative approaches offer advantages in generalization to out-of-distribution data [167] and in maintaining independence from the measurement process [168]. Jalal et al. [169], Song et al. [168], Chung and Ye [170], Chung et al. [167], Cao et al. [171] have developed score-based medical image reconstruction methods that outperform other supervised learning techniques. In contrast, Peng et al. [172], Xie and Li [173] have introduced methods based on denoising diffusion probabilistic models, which demonstrate superior accuracy due to their flexibility in controlling the noise distribution. Cao et al. [171] suggest performing denoising in the measurement space, ensuring consistency between the reconstructed image and acquired data.

Fluid System. Computational fluid dynamics (CFD) plays a crucial role in engineering and scientific applications. Despite the known governing equations of gas or liquid flows, traditional numerical methods face significant computational challenges, particularly in scenarios involving fine-grained simulations or high-Reynolds numbers. Diffusion models have recently gained popularity for refining coarse-grained flow field data or generating high-fidelity flow data due to their ability to capture the chaotic and stochastic nature of turbulence. Shu et al. [174] adopt a physics-informed neural network to learn the denoising process, which denoises the perturbed fluid field data to generate high-fidelity data. The trained model can reconstruct high-fidelity flow data from low-fidelity or sparsely measured data without retraining. Hu et al. [175] further demonstrates the power of diffusion models to generate high-fidelity flow fields in complex geometric scenarios, employing a U-Net neural network and incorporating obstacle geometry as prompts.

1053 **Weather System.** Global weather forecasting is one of the most crucial problems in the weather
1054 system. Traditionally, weather forecasting relies on numerical weather prediction (NWP), which
1055 solves atmospheric dynamics models to generate deterministic future weather scenarios. However,
1056 given the inherent uncertainty in weather patterns, it is important not only to predict a single probable
1057 scenario of future weather $\mathbf{x}$ but also to assess the probability of various future outcomes $P(\mathbf{x})$
1058 [122, 176]. Considering the non-equilibrium nature of global weather as a dynamical system, coupled
1059 with the ability of diffusion models to fit and sample from arbitrary distributions, these models
1060 present a promising solution for quantifying forecasting uncertainty. Li et al. [122] propose SEEDS,
1061 which, by conditioning on a few NWP-generated scenarios, recovers additional forecasting scenarios
1062 through a denoising process. Drawing inspiration from the similarity between weather data and
1063 video, they use a vision Transformer to model score functions. Price et al. [176] introduce GenCast,
1064 which employs a graph transformer designed for spherical meshes to generate future global weather
1065 conditions based on the current state. This approach significantly outperforms existing NWP models
1066 in both efficiency and accuracy for the first time. Their work demonstrates that diffusion-based
1067 methods can address the blurring issues commonly found in other deterministic machine learning
1068 models, underscoring the advantages of diffusion models in this context.

1069 **Astronomical System.** Human understanding of astronomical systems is deeply reliant on astro-
1070 nomical imaging. These images, which project the three-dimensional universe onto two-dimensional
1071 planes, are inevitably affected by background starlight in addition to the celestial objects being
1072 observed. This interference can be likened to the introduction of noise into an image, establishing a
1073 natural connection between diffusion models and astronomical image processing. Drozdova et al.
1074 [177] and Sortino et al. [178] propose using the diffusion model for astronomical image denoising
1075 and synthesis, showing more efficient and effective performance than other methods.

1076 **NeurIPS Paper Checklist**

1077 **1. Claims**

1078 Question: Do the main claims made in the abstract and introduction accurately reflect the
1079 paper's contributions and scope?

1080 Answer: [\[Yes\]](#)

1081 Justification: The relevant information can be founded the abstract, Section 1, 3, and 4.

1082 Guidelines:

- 1083 • The answer NA means that the abstract and introduction do not include the claims
1084 made in the paper.
- 1085 • The abstract and/or introduction should clearly state the claims made, including the
1086 contributions made in the paper and important assumptions and limitations. A No or
1087 NA answer to this question will not be perceived well by the reviewers.
- 1088 • The claims made should match theoretical and experimental results, and reflect how
1089 much the results can be expected to generalize to other settings.
- 1090 • It is fine to include aspirational goals as motivation as long as it is clear that these goals
1091 are not attained by the paper.

1092 **2. Limitations**

1093 Question: Does the paper discuss the limitations of the work performed by the authors?

1094 Answer: [\[NA\]](#)

1095 Justification: This paper is a position paper without directly providing new methods.

1096 Guidelines:

- 1097 • The answer NA means that the paper has no limitation while the answer No means that
1098 the paper has limitations, but those are not discussed in the paper.
- 1099 • The authors are encouraged to create a separate "Limitations" section in their paper.
- 1100 • The paper should point out any strong assumptions and how robust the results are to
1101 violations of these assumptions (e.g., independence assumptions, noiseless settings,
1102 model well-specification, asymptotic approximations only holding locally). The authors
1103 should reflect on how these assumptions might be violated in practice and what the
1104 implications would be.
- 1105 • The authors should reflect on the scope of the claims made, e.g., if the approach was
1106 only tested on a few datasets or with a few runs. In general, empirical results often
1107 depend on implicit assumptions, which should be articulated.
- 1108 • The authors should reflect on the factors that influence the performance of the approach.
1109 For example, a facial recognition algorithm may perform poorly when image resolution
1110 is low or images are taken in low lighting. Or a speech-to-text system might not be
1111 used reliably to provide closed captions for online lectures because it fails to handle
1112 technical jargon.
- 1113 • The authors should discuss the computational efficiency of the proposed algorithms
1114 and how they scale with dataset size.
- 1115 • If applicable, the authors should discuss possible limitations of their approach to
1116 address problems of privacy and fairness.
- 1117 • While the authors might fear that complete honesty about limitations might be used by
1118 reviewers as grounds for rejection, a worse outcome might be that reviewers discover
1119 limitations that aren't acknowledged in the paper. The authors should use their best
1120 judgment and recognize that individual actions in favor of transparency play an impor-
1121 tant role in developing norms that preserve the integrity of the community. Reviewers
1122 will be specifically instructed to not penalize honesty concerning limitations.

1123 **3. Theory assumptions and proofs**

1124 Question: For each theoretical result, does the paper provide the full set of assumptions and
1125 a complete (and correct) proof?

1126 Answer: [\[Yes\]](#)

Justification: The relevant information can be founded in Appendices A, B, and C.

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: The relevant information can be founded in Appendix D.

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: Code and data can be found in anonymous link in Section 3.

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, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: The relevant information can be founded in Appendix D.

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: The relevant information can be founded in Appendix D.

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.

1241 8. Experiments compute resources

1242 Question: For each experiment, does the paper provide sufficient information on the com-
1243 puter resources (type of compute workers, memory, time of execution) needed to reproduce
1244 the experiments?

1245 Answer: [Yes]

1246 Justification: The relevant information can be found in Appendix D.

1247 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).

1256 9. Code of ethics

1257 Question: Does the research conducted in the paper conform, in every respect, with the
1258 NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

1259 Answer: [Yes]

1260 Justification: The authors have read and abide by the NeurIPS Code of Ethics.

1261 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 consid-
1266 eration due to laws or regulations in their jurisdiction).

1267 10. Broader impacts

1268 Question: Does the paper discuss both potential positive societal impacts and negative
1269 societal impacts of the work performed?

1270 Answer: [NA]

1271 Justification: This paper is a position paper without direct paths for positive or negative
1272 societal impacts.

1273 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 is a position paper that poses no such 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: [NA]

Justification: This paper does not use existing assets.

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: [NA]

Justification: This paper has no new assets.

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 work 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 work 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

1384 Question: Does the paper describe the usage of LLMs if it is an important, original, or
1385 non-standard component of the core methods in this research? Note that if the LLM is used
1386 only for writing, editing, or formatting purposes and does not impact the core methodology,
1387 scientific rigorousness, or originality of the research, declaration is not required.

1388 Answer: [NA]

1389 Justification: The core method development in this research does not involve LLMs as any
1390 important, original, or non-standard components.

1391 Guidelines:

- 1392 • The answer NA means that the core method development in this research does not
1393 involve LLMs as any important, original, or non-standard components.
- 1394 • Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>)
1395 for what should or should not be described.