SAMPLING MEETS INTERACTION: SEQUENTIALLY-CONTROLLED MULTI-PARTICLE FLOW-MAPS FOR EFFICIENT INFERENCE-TIME SEARCH

Anonymous authors
Paper under double-blind review

ABSTRACT

While generative models have enabled training-free reward alignment, existing particle-based methods are fundamentally constrained by their propensity to local exploration within narrow regions of the underlying distribution, severely restricting sample diversity. This limitation becomes especially acute under tight reward-feedback budgets, where effective search demands broad, strategic exploration to uncover high-utility regions. To address this, we propose Sequentially-Controlled **I**nteractive **M**ulti-**P**article **F**low-**M**aps (**IMPFM**), a framework for feedback-efficient search. IMPFM progressively transports a group of interactive particles toward the target distribution, maintaining the broad coverage essential for heterogeneous preference alignment. IMPFM leverages a principled and efficient posterior sample-sharing mechanism across particles powered by flow maps. By correcting individual particle drift with the collective value gradient from the entire ensemble's posterior samples at each correction step, the framework maximizes sample utility to enable global exploration while actively mitigating reward over-optimization, typical of standard control frameworks. Paired with a principled exploration-exploitation reweighting mechanism involving multi-particle interaction, this sequentially corrected multi-particle dynamics explicitly preserves structural diversity and overcomes the weight degeneracy inherent to standard Sequential Monte Carlo (SMC) samplers. Crucially, we prove that the resulting sampling framework yields a multi-particle interaction-aware Feynman-Kac corrector that progressively steers the multi-particle system toward a KL-tilted target distribution, facilitating global exploration and preventing mode collapse. Extensive empirical evaluations and ablations across diverse search and alignment tasks confirm the efficacy of IMPFM over existing baselines.

1 INTRODUCTION

Across diverse scientific and engineering disciplines, the pace of discovery is fundamentally constrained by the high cost of evaluation. This bottleneck gives rise to the ubiquitous challenge of feedback-efficient search, where finding an optimal solution requires iteratively proposing candidates and adapting based on sequential feedback. Whether designing life-saving therapeutics or optimizing interactive visual recommendation systems, these real-world applications often rely on costly physical experiments or human engagement. Consequently, minimizing the number of such interactions is a central requirement. Accelerating breakthroughs in these domains demands a robust sampling framework capable of strategic, global exploration across complex, high-dimensional spaces—such as natural images.

Existing methods remain fundamentally ill-equipped for this challenge. For instance, online RL fine-tuning of diffusion models (Uehara et al., 2024) relies on a mode-seeking KL-divergence objective that inherently fails to capture diverse, high-utility regions. Furthermore, these approaches depend on online-trained reward models, whose early-stage biases frequently misdirect fine-tuning and degrade sample quality. Conversely, recent Sequential Monte Carlo (SMC) inference-time search methods (Singhal et al., 2025; Kim et al., 2025) avoid fine-tuning by leveraging fixed generative priors. Yet, they suffer from severe weight degeneracy, leading to proposal collapse and diminished diversity (Lee et al., 2025). Moreover, existing SMC-based samplers rely rigidly on prior generative

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

dynamics, failing to actively correct search trajectories using sequential feedback. This passive exploration inevitably yields suboptimal performance. Recently, Feynman-Kac correctors (Skreta et al., 2025) (FKCs) have emerged as a powerful alternative—progressively correcting dynamics at every resampling step—making them a highly promising paradigm for rapid, feedback-driven adaptation. However, these approaches rely on reward values or gradients at every resampling step; in practice, these signals are unavailable and must be approximated, introducing systematic bias into the sampler and degrading sample quality. Crucially, standard FKC samplers transport particles in strict isolation. This lack of inter-particle interaction severely bottlenecks broad strategic exploration. To overcome these bottlenecks, we introduce *Interactive Multi-Particle Flow-Maps (IMPFM)*, a principled, feedback-efficient sampling framework that ensures diverse coverage across high-reward regions. Figure 1 provides a conceptual overview of IMPFM.

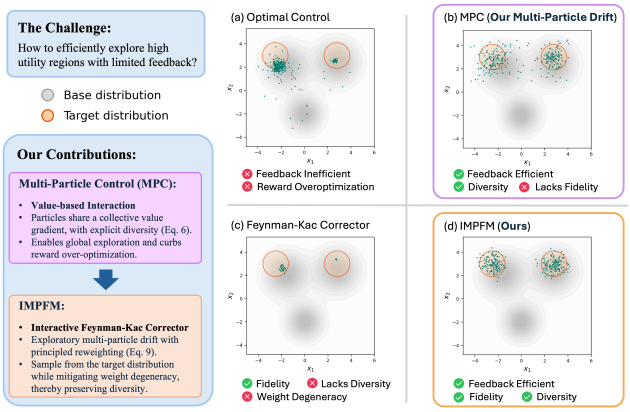


Figure 1: Conceptual Overview of IMPFM.

Key Idea behind IMPFM

IMPFM is built on the premise that search thrives not in isolation, but through collective intelligence. It recognizes that collaboration is the ultimate engine for expanding information diversity and unlocking truly sample-efficient search.

At the heart of IMPFM is a value-based particle interaction mechanism that drives a dual-force exploration dynamic, combining an attractive pull toward high-utility regions with a repulsive push that prevents mode collapse. Powered by Flow Maps, IMPFM enables each particle to exploit the ensemble's collective posterior to optimally correct its drift at every step, and this information reuse mechanism facilitates highly informed, feedback-efficient search. IMPFM's power is substantially enhanced when paired with a principled reweighting mechanism at each resampling step. This reweighting encodes rich value-based multi-particle interactions and enforces a dual-force dynamic—attracting particles to high-density targets while simultaneously exerting a repulsive push to prevent mode collapse and tame reward overoptimization, the classic pitfall of reweighting-based samplers. Critically, we show that IMPFM induces interaction-aware Feynman–Kac correctors that rapidly correct the search dynamics based on sequential feedback and drive the particle system toward the KL-tilted marginal distribution, with the true target as its reference measure. These findings show that, *rather than treating samples independently, it is more powerful to reason about the particle ensemble as a whole and minimize a distributional distance. This ensemble-level view provides a broader picture, promotes global convergence, and yields substantially wider coverage.*

In summary, we make the following contributions:

- **Interactive FKC Sampler**: We introduce IMPFM, a Multi-Particle Interaction-aware Feynman–Kac Corrector for feedback-efficient search.
- **Flow-Map-Driven Information Reuse:** IMPFM leverages a collective posterior-sharing mechanism powered by Flow Maps, maximizing the utility of every drawn sample, transforming isolated particle updates into a highly informed, globally coordinated search.
- **Rapid Adaptation and Coverage**: Driven by value-based dual-force (attractive-repulsive) exploration dynamics, and by coupling collaborative drift correction with dynamic reweighting, IMPFM enables expansive search-space coverage.
- **Rigorous Empirical Validation**: We validate the effectiveness of each component of IMPFM through comprehensive quantitative and qualitative ablation studies.

2 Related Work

Multi-Particle Interaction based Transport Stein Variational Gradient Descent (SVGD) (Liu & Wang, 2016; Liu, 2017) transports interacting particles to match a target distribution. However, it relies on iterative score function evaluations for each particle; approximating these scores introduces biases at every iteration that severely limit its application in high-dimensional settings. Furthermore, SVGD is not designed for reward-tilted distributions, which would additionally require value gradient evaluation of each constituent particle at every step. In stark contrast, our approach entirely bypasses score computation. We introduce a purely value-based particle interaction mechanism—efficiently computed via flow maps—that uses particle interaction to augment exploitative particle drift with broad exploration. While our interactive dynamics target a Boltzmann distribution, we seamlessly recover the true reward-tilted distribution through a principled reweighting mechanism.

Inference-time Scaling for Reward Alignment: Inference-time scaling has garnered significant attention recently, with particle-based SMC emerging as a highly versatile approach that maintains a population of samples to approximate the target distribution. SMC (Del Moral et al., 2006; Phillips et al., 2024; Cardoso et al., 2023; Singhal et al., 2025; Kim et al., 2025) leverages potential functions, which usually approximate the soft value function, to assign weights to particles and resample them at every step. While numerous adaptations have been developed specifically for diffusion model alignment (Dou & Song, 2024; Cardoso et al., 2023; Kim et al., 2025; Trippe et al., 2022; Wu et al., 2023), practical implementations face significant hurdles. Although SMC provides theoretical guarantees of exact sampling in the infinite-particle limit, practical implementations are hindered by weight degeneracy and inaccurate value estimates. These issues often induce mode collapse and compromise sample diversity. Recently, (Holderrieth et al., 2026) proposed Diamond Maps, which utilizes sufficient statistics (Holderrieth et al., 2025) to emulate DDPM-style stochasticity within an ODE-based flow model, thereby facilitating an SMC-like sampling procedure. However, this method inherits the same fundamental limitations as traditional SMC. Recently, several FKC techniques (Skreta et al., 2025; Lee et al., 2025) have been introduced to alleviate particle weight degeneracy and debias particle-based guidance. However, these methods struggle in high-dimensional domains like images. Because they rely on reward gradients at every resampling step—which are typically unavailable and require approximation—they inadvertently inject systematic bias into the sampling process. Although numerous gradient-free alignment methods avoid computing reward gradients by employing BON sampling (Stiennon et al., 2020; Beirami et al., 1879; Nakano et al., 2022), this generate-and-filter approach introduces substantial computational overhead. Furthermore, because these methods only intervene post-hoc at the final output stage, they cannot actively guide the intermediate generative trajectory or fundamentally alter the underlying dynamics. An extended discussion on Related work is provided in the Appendix M.

3 Method: IMPFM Framework

Background: The cornerstone of feedback-efficient search is a policy capable of efficiently sampling from the look-ahead value-tilted distribution. Formally, under optimal control, the marginal density p_t^* of X_t^* at timestep t satisfies:

$$p_t^*(x) \propto p_t(x) e^{V_t(x)}, \quad \text{where} \quad V_t(x) = \log \mathbb{E}\left[e^{r(X_1)} \mid X_t = x \right] \tag{1}$$

Here, p_t denotes the marginal density associated with the unconditional prior process. This formulation ensures that the terminal marginal distribution, p_1^*, aligns exactly with the target reward-tilted distribution $p_{\text{reward}}(x) \propto p_1(x) e^{r(x)}$. Leveraging Doob's h-transform (Denker et al., 2024; Dai Pra, 1991), the optimal drift b_t^* for a diffusion schedule σ_t can be expressed as:

$$\dot{x}_t^* = \underbrace{b_t(x_t^*) + \frac{\sigma_t^2}{2} \nabla V_t(x_t^*)}_{b_t^*(x_t^*)}, \quad \text{with} \quad \text{Law}(x_t^*) = p_t^*, \quad x_0^* \sim p_0 \tag{2}$$

While b_t is directly accessible from the pre-trained Flow Matching model, the primary algorithmic bottleneck lies in efficiently estimating the value function gradient, $\nabla V_t(x)$. A consistent Monte Carlo estimator of $\nabla V_t(x)$ is:

$$\nabla_{x_t} V_t(x_t) = \nabla_{x_t} \log \left(\frac{1}{N} \sum_{i=1}^{N} \exp\left[r(\hat{x}_1^i) \right] \right), \quad \boxed{\hat{x}_1^i \overset{\text{iid}}{\sim} p_{1|t}(\cdot|x_t)} \tag{3}$$

Hence, the primary challenge boils down to efficiently drawing samples from $p_{1|t}(\cdot|x_t)$ via an ODE-based flow model. Conditioned on x_t, one can define an ODE that transports the prior p_0

directly to the posterior $p_{1|t}(\cdot|x_t)$. The associated drift field, $b_s(\cdot; x_t)$, is derived via standard flow matching (Lipman et al., 2022), with the crucial distinction that the objective targets the conditional posterior $p_{1|t}(\cdot|x_t)$ rather than the marginal data distribution p_1:

$$b_s(x; x_t) = \mathbb{E}\left[\dot{\alpha}_s x_1 + \dot{\beta}_s x_0 \mid x_s = x\right], \quad x_s = \alpha_s x_1 + \beta_s x_0, \quad x_0 \sim \mathcal{N}(0, I_d), \quad x_1 \sim p_{1|t}(\cdot|x_t)$$

Consequently, the probability flow associated with $b_s(\cdot; x_t)$ satisfies:

$$\frac{d}{ds} x_s = b_s(x_s; x_t), \quad x_0 \sim \mathcal{N}(0, I_d) \implies \text{Law}(x_1) = p_{1|t}(\cdot|x_t) \tag{4}$$

Where α_t and β_t denote time-dependent coefficients with boundary conditions $\alpha_0 = \beta_1 = 0$ and $\alpha_1 = \beta_0 = 1$. Under the assumption of a Gaussian prior p_0, the conditional drift b_s admits an analytical derivation (Holderrieth et al., 2025) via a reparameterization of the unconditional drift b_t:

$$b_s(x_s; x_t) = w_1 x_s + w_2 b_{t^*}(S(x_s, x_t)) + w_3 x_t$$

Where, w_1, w_2, and w_3 represent scalar weights. Furthermore, S acts as a linear *sufficient statistic*, while t^* represents a reparameterized time variable:

$$S_{s,t}(x_s, x_t) = \frac{\alpha_s \sigma_t^2 x_s + \alpha_t \sigma_s^2 x_t}{\sigma_t^2 \alpha_s^2 + \alpha_t^2 \sigma_s^2}, \qquad t^*(s,t) = g^{-1}\left(\frac{\sigma_t^2 \sigma_s^2}{\sigma_t^2 \alpha_s^2 + \alpha_t^2 \sigma_s^2}\right), \quad g(t) = \frac{\sigma_t^2}{\alpha_t^2} \tag{5}$$

While this reparameterization renders the conditional drift b_s tractable, integrating the ODE trajectories as defined in 4 to draw posterior samples from $p_{1|t}(\cdot|x_t)$ remains a significant computational bottleneck. To bypass this computational bottleneck, a standard strategy is *teacher-distillation*, which leverages the exact analytical field $b_s(x_s; x_t)$ as the teacher for the distillation process. The student velocity model, $\hat{v}$, is optimized via a joint objective function. The first component is an *instantaneous loss* that minimizes the mean squared error between the student's predicted velocity, $\hat{v}_{s,s}$, and the exact analytical teacher field:

$$\mathcal{L}_{\text{inst}}^{\text{distill}}(\hat{v}) := \int_0^1 \int_0^1 \mathbb{E}||\hat{v}_{s,s}(x; x_t) - b_s(x; x_t)||^2 ds dt$$

This is paired with a *consistency loss* (Boffi et al., 2025) ($\mathcal{L}_{\text{cons}}^{\text{distill}}$) to ensure the student reliably integrates this velocity across arbitrary time intervals, defined mathematically as:

$$\int_0^1 \int_0^u \mathbb{E}\left\|\hat{v}_{s,u}(I_s; I_t') - \text{sg}\left(b_u(\hat{X}_{s,u}(I_s; I_t')); I_t'\right) - (u-s)\partial_u \hat{v}_{s,u}(I_s; I_t')\right\|^2 ds\, du$$

Here, $\hat{X}_{s,u}(\cdot\,; x_t) : \mathbb{R}^d \to \mathbb{R}^d$ denotes the solution operator for the context-dependent ODE defined in 4. This distillation process effectively collapses the iterative, multi-step ODE integration into a single amortized map via $\hat{v}$, *enabling computationally efficient, one-step sampling directly from the conditional posterior* $p_{1|t}(\cdot|x_t)$, thereby streamlining the evaluation of $\nabla V_t(x)$ as defined in 3.

Background on FKC: We present the background on the FKC sampler in the Appendix N.

Proposed Approach: Merely repeatedly simulating the optimal drift b_t^* is inadequate for efficient feedback-driven search: solely value-guided trajectories are inherently exploitative and lack an explicit exploratory mechanism, causing them to miss disjoint, high-utility modes within the target landscape. Furthermore, this approach is highly vulnerable to reward over-optimization unless stabilized by an impractically large ensemble of Monte Carlo posterior samples, forcing an untenable trade-off under strict sampling budgets. In particular, Potaptchik et al. (2026) shows that the distance to the true target distribution is bounded by the inverse of the number of Monte Carlo posterior samples utilized to compute ∇V_t at each drift correction step. To resolve this, motivated by (Liu & Wang, 2016), we propose an interactive multi-particle drift correction mechanism that integrates the value gradients from the posterior samples across an ensemble. This introduces an explicit exploration mechanism to augment the purely exploitative optimal drift, b_t^*. By pooling this global information to calibrate individual particle drift, our framework drives broad exploration and prevents reward over-optimization. Specifically, the dynamics of the i-th particle are steered by the following rule:

$$\dot{x}_t^i = b_t(x_t^i) + \underbrace{\frac{\sigma_t^2}{2} \nabla_{x_t^i} V_t(x_t^i)}_{\textbf{Exploitation}} + \underbrace{\frac{\sigma_t^2}{2} \frac{1}{n-1} \sum_{j=1, j\neq i}^{n} \left[\underbrace{k(x_t^j, x_t^i) \nabla_{x_t^j} V(x_t^j)}_{\text{Pulling towards high utility region}} + \underbrace{\nabla_{x_t^j} k(x_t^j, x_t^i)}_{\text{Repulsive force induce diversity}} \right]}_{\textbf{Exploration}: \text{ Interaction between particles encourage broad coverage of high utility regions}} \tag{6}$$

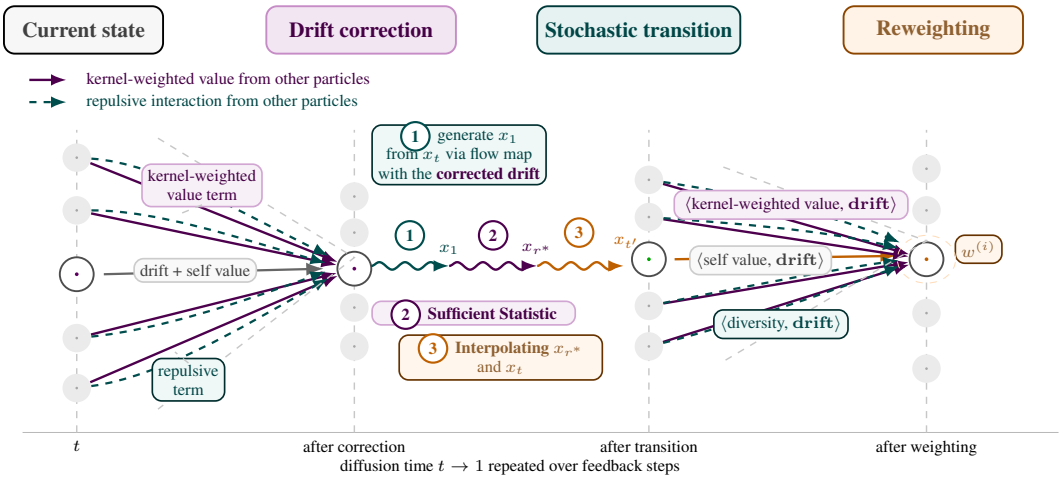


Figure 2: Sampling Mechanism of IMPFM.

Here, n represents the number of particles, $k(x_t^i, x_t^j)$ denotes the Radial Basis Function (RBF) kernel, defined as $k(x_t^i, x_t^j) = \exp\left(-\frac{1}{h}\|x_t^i - x_t^j\|_2^2\right)$, where h is the bandwidth parameter, and $\nabla V_t(x)$ is efficiently computed via Flow-Map following 3. We refer to the resulting framework as Multi-Particle Control (i.e., MPC). In Proposition 4.2 below, we demonstrate that the interplay between the exploitation and exploration terms in Eq. 6 induces a joint steering of the particle dynamics that achieves a steepest descent in the KL divergence between the empirical particle distribution and the Boltzmann distribution, where the value $V(x)$ serves as the energy function.

Proposition 4.2 (Exploration-Augmented Drift Correction via Interactive Particle Dynamics)

Let $q(x)$ denote the empirical distribution induced by the interacting particles, and define the Boltzmann distribution as $p_{\text{boltz}}(x) \propto e^{V(x)}$. The steepest descent direction that minimizes the Kullback-Leibler divergence, i.e., $\nabla_x D_{\text{KL}}(q \parallel p_{\text{boltz}})$ within the unit ball of a Reproducing Kernel Hilbert Space (RKHS), is given by:

$$\mathbb{E}_{x^i \sim q}\left[\nabla_{x^i} V(x^i) + \frac{1}{n-1}\sum_{j=1, j\neq i}^{n}\left[k(x^j, x^i)\nabla_{x^j} V(x^j) + \nabla_{x^j} k(x^j, x^i)\right]\right] \tag{7}$$

Note that the exploration mechanism is driven by a dual-force mechanism. The first component executes a kernel-weighted value-gradient ascent, steering the ensemble toward high-utility manifolds. Complementing this, the second component exerts a repulsive force that prevents particle collapse into local modes, ensuring the ensemble maintains a broad, diverse coverage. Equation 6 also reveals a powerful collaborative mechanism: the drift of the i-th particle is calibrated by the value gradients of its peers. Since computing the value gradient fundamentally requires drawing multiple Monte Carlo samples from a particle's corresponding posterior. Hence, by aggregating Monte Carlo posterior samples across the entire ensemble, the framework effectively pools global information to refine individual trajectories. This collective sharing transforms the estimation of ∇V_t from a local, high-variance task into a robust, ensemble-wide operation, ensuring highly informed corrections. However, our objective is to sample from a value-tilted distribution ($p_1(x)e^{r(x)}$) rather than a standard Boltzmann distribution ($e^{V(x)}$). Because the Boltzmann distribution lacks explicit constraints on distributional coverage, relying on it risks sampling outside the true data manifold. To resolve this, we introduce a reweighting mechanism that guarantees our interactive multi-particle system converges strictly to the desired target distribution while preserving enhanced exploration capabilities of MPC.

To construct a principled reweighting mechanism, we begin by recognizing that while highly efficient, the deterministic nature of ODE sampling inherently restricts exploration, necessitating stochastic transitions for effective search. To achieve this, we first sample standard noise $\epsilon \sim \mathcal{N}(0, I)$ and push it forward through the distilled flow map $\hat{v}(\epsilon; x_t)$. By definition, this directly yields exact samples from

the true posterior, $x_1 \sim p_{1|t}(x_1|x_t)$. With the trajectory's boundary points (ϵ and x_1) firmly anchored, we induce a DDPM-style dynamic duration transition via the Sufficient Statistic (Holderrieth et al., 2025). Specifically, we perform the following steps:

$$\epsilon \xrightarrow{\hat{v}(\epsilon;x_t)} x_1 \sim p_{1|t}(x_1|x_t) \xrightarrow{\text{Interp.}} x_{r^*} = \bar{\alpha}_{r^*}x_1 + \bar{\sigma}_{r^*}\varepsilon_1 \xrightarrow{\text{Trans.}} x_{t'} = \alpha_{t'}\frac{\bar{\alpha}_{r^*}\sigma_t^2 x_{r^*} + \alpha_t\bar{\sigma}_{r^*}^2 x_t}{\bar{\alpha}_{r^*}^2\sigma_t^2 + \alpha_t^2\bar{\sigma}_{r^*}^2},$$
$$r^*(t,t') = g^{-1}\left(\frac{g(t)g(t')}{g(t)-g(t')}\right), \quad g(t) = \frac{\sigma_t^2}{\alpha_t^2}, \quad \varepsilon \sim \mathcal{N}(0,I), \quad \varepsilon_1 \sim \mathcal{N}(0,I). \tag{8}$$

$\bar{\alpha}$ and $\bar{\sigma}^2$ are defined in the Appendix L. Here, the sampled x_1 serves as fixed boundary points to compute the stochastic interpolant x_{r^*}, which acts as the essential foundation for constructing the final DDPM-style transition sample $x_{t'}$.

Armed with DDPM-style transitions, the interactive drift-correction mechanism naturally facilitates a FKC-style reweighting scheme, further amplifying the sampler's capability. While standard value-based reweighting greedily preserves high-reward trajectories at the expense of diversity—inevitably leading to weight degeneracy and mode collapse—we introduce a principled, interaction-aware alternative. By leveraging ensemble-wide multi-particle dynamics, our proposed reweighting mechanism actively preserves the structural diversity of high-utility samples, encouraging broad exploration and global convergence. Specifically, we introduce the following scoring function to weight each particle

(Proposed Particle Reweighting Mechanism Involving Multi-particle Interaction)

$$\underbrace{\langle\nabla_{x_t^i}V(x_t^i), \boldsymbol{b_t(x_t^i)}\rangle + \frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\Big[\langle k(x_t^j,x_t^i)\nabla_{x_t^j}V(x_t^j), \boldsymbol{b_t(x_t^i)}\rangle + \langle\nabla_{x_t^j}k(x_t^j,x_t^i), \boldsymbol{b_t(x_t^i)}\rangle\Big]}_{\textbf{weight of particle } i \textbf{ (i.e. } w_t^{(i)}\textbf{)}} \tag{9}$$

Intuitively, 9 prevents off-manifold degradation by projecting the value and diversity gradients onto the prior model dynamics $b_t(x_t^i)$. A particle is weighted highly only if its pursuit of high reward—both locally and via the ensemble—aligns with the base model dynamics, preventing it from straying into low-likelihood, off-manifold regions. Similarly, the repulsive interaction term guarantees that the drive for sample diversity does not compromise sample quality, as exploratory trajectories are strictly penalized if they violate the prior model dynamics.

To maintain sampling efficiency, we trigger a standard resampling mechanism whenever the effective sample size (ESS) falls below a defined threshold: $\text{ESS}_t^n = \frac{\left(\sum_{i=1}^n w_t^{(i)}\right)^2}{\sum_{i=1}^n \left(w_t^{(i)}\right)^2} \in [1,n]$. Crucially, this resampling procedure leverages importance sampling to ensure the resulting samples remain asymptotically unbiased. Next, we demonstrate that coupling the interactive drift correction (Eq. 6) with our reweighting mechanism (Eq. 9) yields an *interaction-aware Feynman-Kac corrector*, as formalized in Proposition 4.3.

Proposition 4.3 (IMPFM Yields Feynman-Kac Corrector to KL-Tilted Target)

Consider a Flow model with drift $b_t(x)$ trained to sample from $q_t(x)$. Sampling from the KL-tilted marginals $p_t^{\text{kl}}(x) \propto q_t(x)\exp\Big(-D_{\text{KL}}\big(q_t(x) \parallel p_{\text{boltz}}(x)\big)\Big)$ is performed by an weighted SDE with an equivalent corresponding weighted ODE:

$$dx_t^i = b_t(x_t^i)\,dt + \frac{\sigma_t^2}{2}\frac{1}{n}\sum_{j=1}^{n}\Big[k(x_t^j,x_t^i)\nabla_{x_t^j}V(x_t^j)\,dt + \nabla_{x_t^j}k(x_t^j,x_t^i)\,dt\Big] \tag{10}$$
$$dw_t^i = \frac{1}{n}\sum_{j=1}^{n}\Big[\langle k(x_t^j,x_t^i)\nabla_{x_t^j}V(x_t^j), b_t(x_t^i)\rangle\,dt + \langle\nabla_{x_t^j}k(x_t^j,x_t^i), b_t(x_t^i)\rangle\,dt\Big] \tag{11}$$

Thus, as shown in Proposition 4.3, our framework jointly steers the particle ensemble toward the true reward-tilted distribution while maintaining high likelihood under the prior to preserve intrinsic sample quality. By leveraging multi-particle interactions to explicitly minimize the KL divergence to the target, this holistic approach fundamentally circumvents the severe vulnerabilities of reward

over-optimization and mode collapse inherent in isolated trajectory optimization. We term our framework Sequentially Controlled Interactive Multi-Particle Flow-Maps (IMPFM) (See Fig. 2).

To broaden the applicability of the IMPFM to settings lacking an off-the-shelf stochastic flow map, we seamlessly adapt our approach using a standard deterministic flow matching model—paired with a consistency model enabling a single-step sample (i.e., x_1) generation from any x_t. However, the inherent determinism of consistency models precludes sampling multiple $x_1^{(i)}$ from the posterior $p_{1|t}(\cdot|x_t)$, a critical requirement for estimating the value gradient ∇V_t and key to computing the corrected dynamics (Eq. 6) and reweighting scores (Eq. 9). To bypass this limitation, we introduce an *iterative posterior sampling scheme: we alternate between predicting x_1 from x_t via the consistency model, and injecting noise via a stochastic interpolant to produce a perturbed x_t. Repeating this sequence effectively yields valid samples (i.e., $x_1^{(1)}, x_1^{(2)}, \dots x_1^{(K)},$) from the desired posterior*, as depicted below (where $\epsilon_1, \epsilon_2, \dots, \epsilon_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I)$):

$$x_t^{(0)} \xrightarrow{\text{Consistency}} x_1^{(1)} \xrightarrow{x_t^{(1)}=\alpha_t x_1^{(1)}+\sigma_t\epsilon_1} x_t^{(1)} \xrightarrow{\text{Consistency}} x_1^{(2)} \xrightarrow{x_t^{(2)}=\alpha_t x_1^{(2)}+\sigma_t\epsilon_2} x_t^{(2)} \cdots \xrightarrow{\text{Repeat}} x_1^{(K)} \tag{12}$$

4 Experiments and Results

We evaluate IMPFM across two settings. In the *feedback-efficient search* setting, a target-class-conditioned prior operates without a priori knowledge of the fine-grained target prompt, relying entirely on iterative feedback to progressively steer generated samples toward the target specification revealed only through sequential interaction. In the *feedback-driven alignment* setting, a prior model, conditioned on the target prompt, leverages sequential feedback to iteratively refine its outputs toward satisfying complex, fine-grained constraints — including compositional and quantitative reasoning. We benchmark IMPFM against a diverse set of competitive, state-of-the-art baselines. These include FKC (Skreta et al., 2025), Feynman-Kac Steering (FKS) (Singhal et al., 2025), the SMC-based approach DAS (Kim et al., 2025), Best-of-N, and Meta-Flow-Maps with gradient-based estimator (MFM-G, hereafter referred to as MFM) (Potaptchik et al., 2026). To simulate online feedback, we employ widely adopted reward models, i.e., VQA score measured with CLIP-FlanT5 (Lin et al., 2024), and reward score using InstructBLIP (Dai et al., 2023), ImageReward (Xu et al., 2023), and Pick-Score (Kirstain et al., 2023). Sample diversity is evaluated using LPIPS (Kim et al., 2025) score.

Search Experimental Setting: To evaluate the search capabilities of IMPFM, we utilize ImageNet-1K. For the prior, we employ a distilled flow map model optimized jointly using the $\mathcal{L}_{\text{inst}}^{\text{distill}}$ and $\mathcal{L}_{\text{const}}^{\text{distill}}$. We evaluate on 200 randomly selected target classes, revealing their fine-grained specifications strictly via iterative feedback (see Appendix G for example class and prompt details).

Search Result: Fig. 4 reports the VQA scores, reward scores, and diversity evaluated across different feedback budgets ($\mathcal{B}$). Experimental results demonstrate that *IMPFM achieves higher reward scores with considerably fewer feedback than competing approaches*. This highlights its efficacy for feedback-driven search, a setting where maximizing reward while ensuring sample efficiency is paramount. IMPFM achieves superior sample efficiency through a highly effective information-sharing mechanism. Unlike prior approaches, the posterior samples generated by any single particle are shared across the entire ensemble for drift correction and reweighting, maximizing information utilization and significantly boosting the framework's overall sample efficiency.

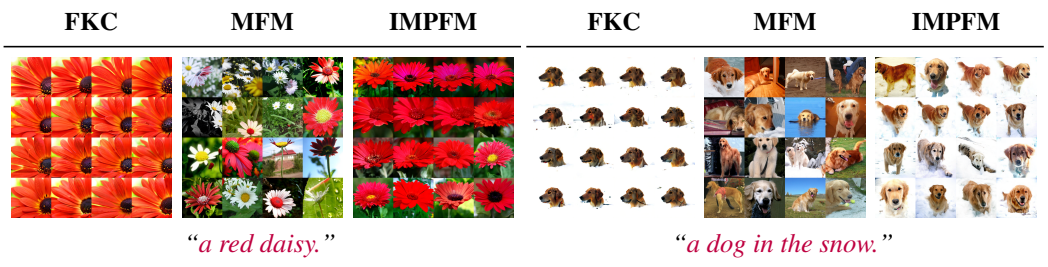


Figure 3: Search Visualizations.

Notably, IMPFM demonstrates superior resistance to reward over-optimization. While both IMPFM and the MFM baseline utilize ImageReward to steer the dynamics, IMPFM consistently outperforms the baselines across held-out evaluation reward models, e.g., Pick-Score (Kirstain et al., 2023). We

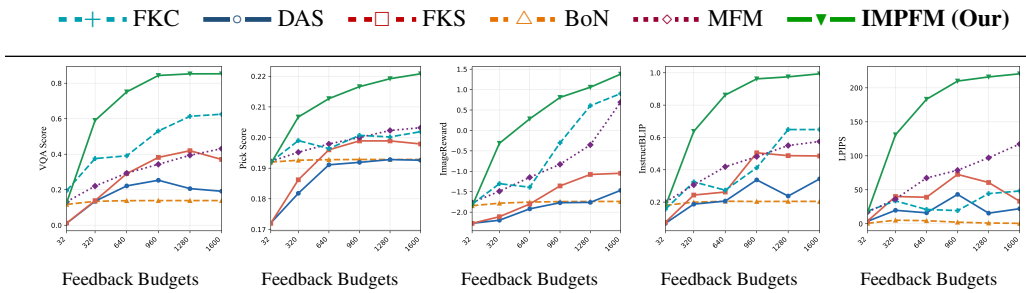
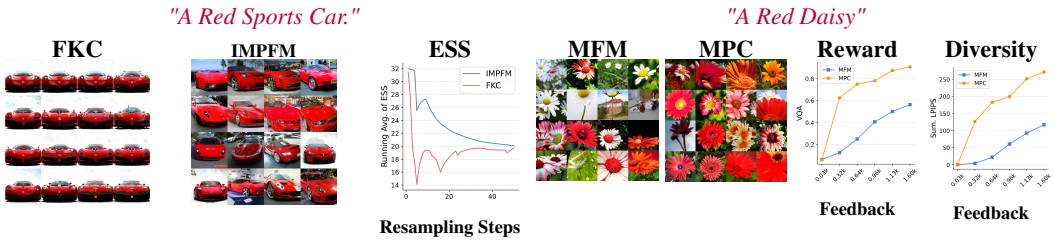


Figure 4: Search Performance Analysis of Competitive Approaches on ImageNet Target Classes.

attribute this robustness to our ensemble-based reward guidance mechanism. Rather than evaluating each particle against local reward signals in isolation, IMPFM aggregates reward feedback across the full ensemble's posterior samples. This global aggregation effectively smooths the reward landscape, establishing a broader consensus for the guidance that prevents individual particles from collapsing into the narrow, high-reward local maxima that typically emerge under purely local evaluation. Crucially, IMPFM unlocks a striking increase in semantic diversity among the generated samples, as reflected in the LPIPS score. This underscores the value of our explicit, diversity-inducing repulsive force—demonstrating that actively enforcing particle separation for reweighting and drift correction is key for diversity. We visualize the generated samples in Fig. 3 (See Appendix H for more viz).



(a) Impact of Particle Interaction on FKC.

(b) Impact of Particle Interaction on Optimal Control.

Emergent Stability: Eradicating Weight Degeneracy via Particle Interaction Weight degeneracy remains a fundamental bottleneck in traditional SMC/FKC frameworks. To isolate the contribution of particle interaction, we conduct two ablation tests. First, we remove the interaction term from both the drift correction (Eq. 6) and reweighting (Eq. 9), reducing the sampler to a standard reward-tilted FKC. Second, we take a purely value-guided optimal control sampler — MFM — and equip it with our proposed interactive drift correction mechanism (Eq. 6), forming a variant we refer to as MPC, to assess the performance gains attributable to interaction alone. As shown in Fig. 5a, the interaction-aware formulation in both drift correction and reweighting is critical: it promotes substantially greater sample diversity, sustains a persistently high ESS across all correction steps— failure modes endemic to the FKC. Similarly, comparative results between MFM and MPC, as depicted in Fig. 5b, confirm that particle interaction-based drift correction (Eq. 6) is key to overcoming mode collapse and reward over-optimization even within a purely value-based optimal control setting, as evidenced by richer sample diversity and consistently higher VQA scores. Additional analysis is in the Appendix F, D.

Importance of Reweighting Mechanism We isolate the effect of the reweighting step by comparing the full IMPFM framework against an ablated variant, referred to as Multi Particle Control (MPC), in which only the reweighting mechanism is removed. As reported in 6, the reweighting step proves critical — it drives faster convergence to higher reward scores (evaluated using the VQA score) across all feedback budgets. As shown in 6, MPC yields some suboptimal samples because it lacks explicit constraints to remain on the true data manifold. Our reweighting mechanism effectively corrects this, underscoring its necessity in the IMPFM framework.

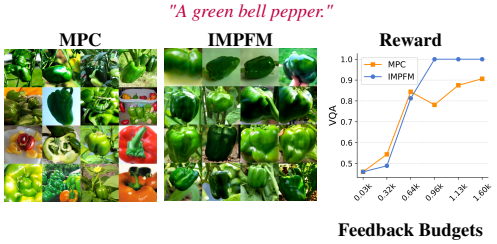


Figure 6: Impact of Reweighting.



Figure 8: Alignment Visualization. (left) Quantity Prompt. (right) Compositional Prompt.

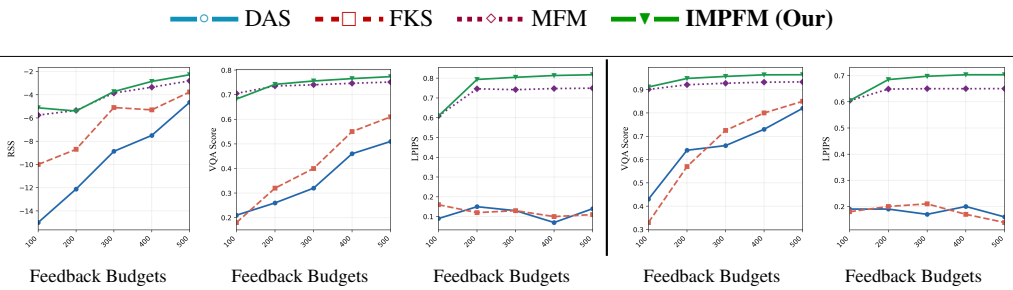


Figure 9: Analysis of Competitive Approaches on (right) Compositional, (left) Quantity-aware Alignment.

The Role of Sufficient Statistic (SS) SS serves a dual purpose: it introduces DDPM-like stochasticity to enable exploration while retaining ODE-based generative fidelity, and *supports arbitrary-duration stochastic transitions essential for budget-efficient search.* A direct comparison across varying feedback budgets (Fig. 7) reveals that the SS-based formulation yields consistently good sample quality — even under severe feedback-budget constraints — confirming its importance for efficient feedback-driven search. Additional analysis on the role of Sufficient Statistic on search is provided in the Appendix E.



Figure 7: Efficacy of SS on IMPFM. (left) $\mathcal{B} = 20$; (right) $\mathcal{B} = 80$

Alignment Experimental Setting: To assess the effectiveness of IMPFM feedback-driven alignment problems, we focus on two challenging tasks. First, for compositional alignment, we evaluate using 200 randomly selected prompts from GenAI-Bench (Jiang et al., 2024) that involve at least three advanced compositional skills. Second, for quantity-aware alignment, we evaluate using 200 prompts randomly sampled from the numeracy category of T2I-CompBench++ (Huang et al., 2025). For the quantity-aware alignment task, the reward is also evaluated as the negative Residual Sum of Squares (RSS) between the target object counts and the detected counts, where detections are obtained using DINO (Liu et al., 2024). We employ an off-the-shelf consistency model (Chen et al., 2025) as the prior and sample from the posterior ($p_{1|t}(\cdot|x_t)$) using the proposed iterative posterior sampling scheme. We provide a detailed derivation (see App. K) of the reparametrization of the noise schedule used in (Chen et al., 2025) to facilitate its integration within the IMPFM framework.

Alignment Result: The Alignment results (Fig. 9) reveal a similar trend. IMPFM converges to high reward scores with significantly fewer feedback interactions than competing approaches, while simultaneously maintaining greater sample diversity. The rapid convergence further underscores the effectiveness of the particle interaction mechanism, which enables maximum information utilization for both drift correction and reweighting. The sustained diversity, in turn, confirms that IMPFM does not suffer from the weight degeneracy issues that commonly afflict practical inference — a property made possible by the explicit repulsive force between particles, which actively discourages collapse and promotes population-level diversity. Qualitative results are visualized in Fig. 17 (See O). Furthermore, these findings underscore the effectiveness of the iterative posterior sampling scheme.

Conclusion In this work, we introduce IMPFM, a principled, interactive Feynman-Kac corrector (FKC) sampler designed for sample-efficient feedback-driven search and alignment. Our rigorous empirical analysis demonstrates its scalability and efficacy, particularly in high-dimensional settings. In the future, we aim to extend the applicability of IMPFM across diverse scientific domains.

AI USE STATEMENT

We used large language models (LLMs) solely as a writing assistant for language polishing and improving clarity of presentation. The LLMs were not involved in research ideation, methodological design, experimental execution, or result analysis. All scientific contributions and substantive writing were carried out by the authors.

REPRODUCIBILITY STATEMENT

To facilitate reproducibility, we have provided full implementation details, hyperparameter settings, and architectural specifications. Our complete source code, including training scripts and evaluation pipelines, is available via the zip file included in the supplementary material. Upon acceptance, we will make all code publicly available to facilitate easy reproduction.

REFERENCES

Ahmad Beirami, Alekh Agarwal, Jonathan Berant, Alex D'Amour, Jacob Eisenstein, Chirag Nagpal, and Ananda Theertha Suresh. Theoretical guarantees on the best-of-n alignment policy. 2024. *URL https://api. semanticscholar. org/CorpusID*, 266741736(3), 1879.

Nicholas M Boffi, Michael S Albergo, and Eric Vanden-Eijnden. How to build a consistency model: Learning flow maps via self-distillation. *arXiv preprint arXiv:2505.18825*, 2025.

Gabriel Cardoso, Yazid Janati El Idrissi, Sylvain Le Corff, and Eric Moulines. Monte carlo guided diffusion for bayesian linear inverse problems. *arXiv preprint arXiv:2308.07983*, 2023.

Junsong Chen, Shuchen Xue, Yuyang Zhao, Jincheng Yu, Sayak Paul, Junyu Chen, Han Cai, Song Han, and Enze Xie. Sana-sprint: One-step diffusion with continuous-time consistency distillation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 16185–16195, 2025.

Wenliang Dai, Junnan Li, Dongxu Li, Anthony Tiong, Junqi Zhao, Weisheng Wang, Boyang Li, Pascale N Fung, and Steven Hoi. Instructblip: Towards general-purpose vision-language models with instruction tuning. *Advances in neural information processing systems*, 36:49250–49267, 2023.

Paolo Dai Pra. A stochastic control approach to reciprocal diffusion processes. *Applied mathematics and Optimization*, 23(1):313–329, 1991.

Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential monte carlo samplers. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 68(3):411–436, 2006.

Alexander Denker, Francisco Vargas, Shreyas Padhy, Kieran Didi, Simon Mathis, Vincent Dutordoir, Riccardo Barbano, Emile Mathieu, Urszula J Komorowska, and Pietro Lio. Deft: Efficient fine-tuning of diffusion models by learning the generalised h-transform. *Advances in Neural Information Processing Systems*, 37:19636–19682, 2024.

Hanze Dong, Wei Xiong, Deepanshu Goyal, Yihan Zhang, Winnie Chow, Rui Pan, Shizhe Diao, Jipeng Zhang, Kashun Shum, and Tong Zhang. Raft: Reward ranked finetuning for generative foundation model alignment. *arXiv preprint arXiv:2304.06767*, 2023.

Zehao Dou and Yang Song. Diffusion posterior sampling for linear inverse problem solving: A filtering perspective. In *The Twelfth International Conference on Learning Representations*, 2024.

Peter Holderrieth, Uriel Singer, Tommi Jaakkola, Ricky TQ Chen, Yaron Lipman, and Brian Karrer. Glass flows: Transition sampling for alignment of flow and diffusion models. *arXiv preprint arXiv:2509.25170*, 2025.

Peter Holderrieth, Douglas Chen, Luca Eyring, Ishin Shah, Giri Anantharaman, Yutong He, Zeynep Akata, Tommi Jaakkola, Nicholas Matthew Boffi, and Max Simchowitz. Diamond maps: Efficient reward alignment via stochastic flow maps. *arXiv preprint arXiv:2602.05993*, 2026.

Kaiyi Huang, Chengqi Duan, Kaiyue Sun, Enze Xie, Zhenguo Li, and Xihui Liu. T2i-compbench++: An enhanced and comprehensive benchmark for compositional text-to-image generation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(5):3563–3579, 2025.

Vineet Jain, Kusha Sareen, Mohammad Pedramfar, and Siamak Ravanbakhsh. Diffusion tree sampling: Scalable inference-time alignment of diffusion models. *arXiv preprint arXiv:2506.20701*, 2025.

Dongfu Jiang, Max Ku, Tianle Li, Yuansheng Ni, Shizhuo Sun, Rongqi Fan, and Wenhu Chen. Genai arena: An open evaluation platform for generative models. *Advances in Neural Information Processing Systems*, 37:79889–79908, 2024.

Sunwoo Kim, Minkyu Kim, and Dongmin Park. Test-time alignment of diffusion models without reward over-optimization. *arXiv preprint arXiv:2501.05803*, 2025.

Yuval Kirstain, Adam Polyak, Uriel Singer, Shahbuland Matiana, Joe Penna, and Omer Levy. Pick-a-pic: An open dataset of user preferences for text-to-image generation. *Advances in neural information processing systems*, 36:36652–36663, 2023.

Cheuk Kit Lee, Paul Jeha, Jes Frellsen, Pietro Lio, Michael Samuel Albergo, and Francisco Vargas. Debiasing guidance for discrete diffusion with sequential monte carlo. *arXiv preprint arXiv:2502.06079*, 2025.

Zhiqiu Lin, Deepak Pathak, Baiqi Li, Jiayao Li, Xide Xia, Graham Neubig, Pengchuan Zhang, and Deva Ramanan. Evaluating text-to-visual generation with image-to-text generation. In *European Conference on Computer Vision*, pp. 366–384. Springer, 2024.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Qiang Liu. Stein variational gradient descent as gradient flow. *Advances in neural information processing systems*, 30, 2017.

Qiang Liu and Dilin Wang. Stein variational gradient descent: A general purpose bayesian inference algorithm. *Advances in neural information processing systems*, 29, 2016.

Qiang Liu, Jason Lee, and Michael Jordan. A kernelized stein discrepancy for goodness-of-fit tests. In *International conference on machine learning*, pp. 276–284. PMLR, 2016.

Shilong Liu, Zhaoyang Zeng, Tianhe Ren, Feng Li, Hao Zhang, Jie Yang, Qing Jiang, Chunyuan Li, Jianwei Yang, Hang Su, et al. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. In *European conference on computer vision*, pp. 38–55. Springer, 2024.

Reiichiro Nakano, Jacob Hilton, Suchir Balaji, Jeff Wu, Long Ouyang, Christina Kim, Christopher Hesse, Shantanu Jain, Vineet Kosaraju, William Saunders, et al. Webgpt: Browser-assisted question-answering with human feedback, 2022. *URL https://arxiv. org/abs/2112.09332*, 35, 2022.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744, 2022.

Angus Phillips, Hai-Dang Dau, Michael John Hutchinson, Valentin De Bortoli, George Deligiannidis, and Arnaud Doucet. Particle denoising diffusion sampler. *arXiv preprint arXiv:2402.06320*, 2024.

Peter Potaptchik, Adhi Saravanan, Abbas Mammadov, Alvaro Prat, Michael S Albergo, and Yee Whye Teh. Meta flow maps enable scalable reward alignment. *arXiv preprint arXiv:2601.14430*, 2026.

Raghav Singhal, Zachary Horvitz, Ryan Teehan, Mengye Ren, Zhou Yu, Kathleen McKeown, and Rajesh Ranganath. A general framework for inference-time scaling and steering of diffusion models. *arXiv preprint arXiv:2501.06848*, 2025.

Marta Skreta, Tara Akhound-Sadegh, Viktor Ohanesian, Roberto Bondesan, Alán Aspuru-Guzik, Arnaud Doucet, Rob Brekelmans, Alexander Tong, and Kirill Neklyudov. Feynman-kac correctors in diffusion: Annealing, guidance, and product of experts. *arXiv preprint arXiv:2503.02819*, 2025.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020a.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020b.

Nisan Stiennon, Long Ouyang, Jeffrey Wu, Daniel Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul F Christiano. Learning to summarize with human feedback. *Advances in neural information processing systems*, 33:3008–3021, 2020.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023.

Brian L Trippe, Jason Yim, Doug Tischer, David Baker, Tamara Broderick, Regina Barzilay, and Tommi Jaakkola. Diffusion probabilistic modeling of protein backbones in 3d for the motif-scaffolding problem. *arXiv preprint arXiv:2206.04119*, 2022.

Masatoshi Uehara, Yulai Zhao, Kevin Black, Ehsan Hajiramezanali, Gabriele Scalia, Nathaniel Lee Diamant, Alex M Tseng, Sergey Levine, and Tommaso Biancalani. Feedback efficient online fine-tuning of diffusion models. *arXiv preprint arXiv:2402.16359*, 2024.

Luhuan Wu, Brian Trippe, Christian Naesseth, David Blei, and John P Cunningham. Practical and asymptotically exact conditional sampling in diffusion models. *Advances in Neural Information Processing Systems*, 36:31372–31403, 2023.

Jiazheng Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, and Yuxiao Dong. Imagereward: Learning and evaluating human preferences for text-to-image generation. *Advances in Neural Information Processing Systems*, 36:15903–15935, 2023.

SAMPLING MEETS INTERACTION: SEQUENTIALLY CONTROLLED MULTI-PARTICLE FLOW-MAPS FOR EFFICIENT INFERENCE-TIME SEARCH(APPENDIX)

APPENDIX TABLE OF CONTENTS

Appendix: Theoretical Derivations **Page**

1 Derivation of Exploration-Augmented Drift Correction via Interactive Particle Dynamics 14
2 IMPFM Yields Fynman-Kac Corector to KL-Tilted Target 15

Appendix: Additional Quantitative Results on Search Task

1 Performance Comparison in Longer Horizon Search Tasks 17

Appendix: Additional Ablation Studies and Qualitative Analysis of IMPFM

1 Additional Analysis on the Impact of the Interaction-Aware Feynman-Kac Corrector 18
2 Visualizing the Impact of Sufficient Statistic 19
3 Additional Qualitative Analysis: Taming Weight Degeneracy Through Particle Interaction-Guided Drift Correction 20

Appendix: Additional Details and Results on Search

1 Details of Class and Prompt Detail in the Search Experiment 21
2 Additional Comparative Visualization on Search Task 21

Appendix: Additional Implementation Details, Background, and Qualitative Results on Alignment Task

1 Details about the Evaluation Dataset for Alignment Tasks 22
2 Implementation Details 22
3 Derivation of Noise Schedule to Adapt IMPFM on Alignment Tasks 22
4 Sufficient Statistic Computation 24
5 Extended Related Work 24
6 Background: Feynman-Kac Correctors 24
7 Additional Visualizations on the Alignment Task 25

SAMPLING MEETS INTERACTION: SEQUENTIALLY CONTROLLED MULTI-PARTICLE FLOW-MAPS FOR EFFICIENT INFERENCE-TIME SEARCH (APPENDIX)

A PROOF OF PROPOSITION 4.2

Proposition 4.1 (Exploration-Augmented Drift Correction via Interactive Particle Dynamics))

Let $q(x)$ denote the empirical distribution induced by the interacting particles, and define the Boltzmann distribution as $p_{\text{boltz}}(x) \propto e^{V(x)}$. The steepest descent direction that minimizes the Kullback-Leibler divergence, i.e., $\nabla_x D_{\text{KL}}(q \parallel p_{\text{boltz}})$ within the unit ball of a Reproducing Kernel Hilbert Space (RKHS) is given by:

$$\mathbb{E}_{x^i \sim q}\left[\nabla_{x^i} V(x^i) + \frac{1}{n-1}\sum_{j=1, j\neq i}^{n}\left[k(x^j, x^i)\nabla_{x^j} V(x^j) + \nabla_{x^j} k(x^j, x^i)\right]\right] \tag{13}$$

Proof. We aim to derive the steepest direction minimizing the KL-Divergence within the unit ball of an RKHS.

Step 1: Setup the Perturbation Let $q(x)$ be the current distribution of particles. We apply a smooth transformation $T(x) = x + \epsilon\phi(x)$, where $\epsilon > 0$ is a small step size and $\phi \in \mathcal{H}^d$ is the velocity field in a Reproducing Kernel Hilbert Space (RKHS) with kernel $k(\cdot, \cdot)$. Let $q_{[\epsilon\phi]}$ denote the pushforward distribution of q under T.

Step 2: Derivative of the KL Divergence We want to find the direction ϕ that maximizes the decrease of $D_{KL}(q_{[\epsilon\phi]} \parallel p_{boltz})$. By the results of Liu et al. (2016), the first variation of the KL divergence at $\epsilon = 0$ evaluates to the expectation of the Stein operator:

$$\frac{d}{d\epsilon} D_{KL}(q_{[\epsilon\phi]} \parallel p_{boltz})\Big|_{\epsilon=0} = -\mathbb{E}_{x\sim q}\big[\text{trace}(\mathcal{A}_{p_{boltz}}\phi(x))\big]$$

where the Stein operator $\mathcal{A}_p$ is defined as:

$$\mathcal{A}_{p_{boltz}}\phi(x) = \phi(x)^\top \nabla_x \log p_{boltz}(x) + \nabla_x \cdot \phi(x)$$

Given $p_{boltz}(x) \propto e^{V(x)}$, we have $\nabla_x \log p_{boltz}(x) = \nabla_x V(x)$. Substituting this in yields:

$$\frac{d}{d\epsilon} D_{KL}\Big|_{\epsilon=0} = -\mathbb{E}_{x\sim q}\big[\phi(x)^\top \nabla_x V(x) + \nabla_x \cdot \phi(x)\big]$$

Step 3: Applying the RKHS Reproducing Property By the reproducing property of the RKHS, any $\phi \in \mathcal{H}^d$ satisfies $\phi(x) = \langle \phi, k(x, \cdot)\rangle_{\mathcal{H}}$. We can pass the linear inner product outside the expectation:

$$\mathbb{E}_{x\sim q}\big[\phi(x)^\top \nabla_x V(x) + \nabla_x \cdot \phi(x)\big] = \big\langle \phi(\cdot),\ \mathbb{E}_{x\sim q}\big[k(x, \cdot)\nabla_x V(x) + \nabla_x k(x, \cdot)\big]\big\rangle_{\mathcal{H}}$$

Step 4: The Steepest Descent Direction To find the steepest descent direction within the unit ball of the RKHS, we solve:

$$\phi^* = \arg\max_{\|\phi\|_{\mathcal{H}} \leq 1}\left(-\frac{d}{d\epsilon} D_{KL}\Big|_{\epsilon=0}\right)$$

By the Cauchy-Schwarz inequality in Hilbert spaces ($\langle \phi, u\rangle \leq \|\phi\|\|u\|$), the linear functional is maximized when ϕ is strictly proportional to the second argument of the inner product. Therefore, the optimal vector field is:

$$\phi^*(\cdot) = \mathbb{E}_{x\sim q}\big[k(x, \cdot)\nabla_x V(x) + \nabla_x k(x, \cdot)\big]$$

Step 5: Empirical Particle Approximation Given n interacting particles $\{x^1, \dots, x^n\}$, evaluating the optimal velocity field ϕ^* at a specific particle x^i using the empirical distribution $q(x) \approx \frac{1}{n}\sum_{j=1}^{n} \delta(x - x^j)$ yields:

$$\phi^*(x^i) = \frac{1}{n}\sum_{j=1}^{n} \left[k(x^j, x^i)\nabla_{x_j} V(x^j) + \nabla_{x_j} k(x^j, x^i)\right]$$

To match the proposed formulation in Proposition 4.1, we isolate the self-interaction term ($j = i$). For standard translation-invariant kernels (like the RBF kernel), $k(x^i, x^i) = 1$ and $\nabla_{x_i} k(x^i, x^i) = 0$. By separating the $j = i$ term and replacing the generic $\frac{1}{n}$ weight of the interaction sum with the unbiased leave-one-out U-statistic estimator $\frac{1}{n-1}$ for the other particles, the update direction at x^i becomes exactly:

$$\nabla_{x_i} V(x^i) + \frac{1}{n-1}\sum_{j\neq i} \left[k(x^j, x^i)\nabla_{x_j} V(x^j) + \nabla_{x_j} k(x^j, x^i)\right]$$

Taking the expectation over all particles $x^i \sim q$ yields the exact expression as stated in Proposition 4.1.

□

B Proof of Proposition 4.3

Proposition 4.3 (IMPFM Yields Fynman-Kac Corector to KL-Tilted Target)

Consider a Flow model with drift $b_t(x)$ trained to sample from $q_t(x)$. Sampling from the KL-tilted marginals $p_t^{\text{kl}}(x) \propto q_t(x) \exp\Big(-D_{\text{KL}}\big(q_t(x) \parallel p_{\text{boltz}}(x)\big)\Big)$ is performed by an weighted SDE with an equivalent corresponding weighted ODE:

$$dx_t^i = b_t(x_t^i)\, dt + \frac{\sigma_t^2}{2}\frac{1}{n}\sum_{j=1}^{n}\Big[k(x_t^j, x_t^i)\nabla_{x_t^j} V(x_t^j)\, dt + \nabla_{x_t^j} k(x_t^j, x_t^i)\, dt\Big] \tag{14}$$

$$dw_t^i = \frac{1}{n}\sum_{j=1}^{n}\Big[\langle k(x_t^j, x_t^i)\nabla_{x_t^j} V(x_t^j), b_t(x_t^i)\rangle\, dt + \langle \nabla_{x_t^j} k(x_t^j, x_t^i), b_t(x_t^i)\rangle\, dt\Big] \tag{15}$$

Proof. We first show the weighted SDE form that corresponds to sampling from $p_t^{kl}(x)$, then we convert the resulting weighted SDE to the corresponding weighted ODE.

Step 1: Assume $f^{kl}(x) = -D_{KL}(q_t(x)||p_{boltz}(x))$. The generation process can be defined as the family of denoising SDEs in the opposite time direction as below:

$$dx_t = \big(-f_t(x_t) + \sigma_t^2 \nabla \log q_t(x_t)\big)\, dt + \sigma_t dW_t \tag{16}$$

The SDE in Eq. 16 can be viewed as the composition of a flow and diffusion terms, where the corresponding Fokker-Planck PDE describes the combined evolution:

$$\frac{\partial q_t(x)}{\partial t} = -\underbrace{\langle \nabla, q_t(x) v_t(x)\rangle}_{\text{Continuity Eqn.}} + \underbrace{\frac{\sigma_t^2}{2}\Delta q_t(x)}_{\text{Diffusion Eqn.}}. \tag{17}$$

We want to find the PDE for the reward-tilted density:

$$\hat{p_t}^{kl}(x) = \frac{q_t(x)\exp(\beta_t f^{kl}(x))}{\int dx\, q_t(x)\exp(\beta_t f^{kl}(x))}. \tag{18}$$

By applying the logarithm and then taking the partial derivative with respect to t, we get:

$$\frac{\partial}{\partial t}\log \hat{p_t}^{kl}(x) = \frac{\partial}{\partial t}\log q_t(x) + \frac{\partial \beta_t}{\partial t} f^{kl}(x) - \int dx\, \hat{p_t}^{kl}(x)\left[\frac{\partial}{\partial t}\log q_t(x) + \frac{\partial \beta_t}{\partial t} f^{kl}(x)\right]. \tag{19}$$

For the first term, we have:

$$\begin{aligned}\frac{\partial}{\partial t}\log q_t(x) &= -\langle\nabla, v_t(x)\rangle - \langle\nabla\log q_t(x), v_t(x)\rangle + \frac{\sigma_t^2}{2}\Delta\log q_t(x) + \frac{\sigma_t^2}{2}\|\nabla\log q_t(x)\|^2 \\ &= -\langle\nabla, v_t(x)\rangle - \langle\nabla\log \hat{p_t}^{kl}(x), v_t(x)\rangle + \frac{\sigma_t^2}{2}\Delta\log \hat{p_t}^{kl}(x) + \frac{\sigma_t^2}{2}\|\nabla\log \hat{p_t}^{kl}(x)\|^2 \\ &\quad + \left\langle \beta_t\nabla f^{kl}(x), v_t(x) - \sigma_t^2\nabla\log q_t(x) - \frac{\sigma_t^2}{2}\beta_t\nabla f^{kl}(x)\right\rangle - \beta_t\frac{\sigma_t^2}{2}\Delta f^{kl}(x). \quad (20)\end{aligned}$$

The last equality is obtained using the fact that

$$\nabla\log q_t(x) = \nabla\log \hat{p_t}^{kl}(x) - \beta_t\nabla f^{kl}(x)$$

and

$$\Delta\log q_t(x) = \Delta\log \hat{p_t}^{kl}(x) - \beta_t\Delta f^{kl}(x)$$

Both of which follow from 18.

Thus, we have:

$$\frac{\partial \hat{p_t}^{kl}(x)}{\partial t} = -\langle\nabla, \hat{p_t}^{kl}(x)v_t(x)\rangle + \frac{\sigma_t^2}{2}\Delta\hat{p_t}^{kl}(x) + \hat{p_t}^{kl}(x)\big(g_t(x) - \mathbb{E}_{\hat{p_t}^{kl}(x)}[g_t(x)]\big), \quad (21)$$

$$g_t(x) = \left\langle \beta_t\nabla f^{kl}(x), v_t(x) - \sigma_t^2\nabla\log q_t(x) - \frac{\sigma_t^2}{2}\beta_t\nabla f^{kl}(x)\right\rangle - \beta_t\frac{\sigma_t^2}{2}\Delta f^{kl}(x) + \frac{\partial\beta_t}{\partial t}f^{kl}(x). \quad (22)$$

Furthermore, we can add the gradient of the tilted function as an additional drift term $a\nabla f^{kl}(x)$, i.e.,

$$\frac{\partial \hat{p_t}^{kl}(x)}{\partial t} = -\langle\nabla, \hat{p_t}^{kl}(x)(v_t(x) + a\nabla f^{kl}(x))\rangle + \frac{\sigma_t^2}{2}\Delta\hat{p_t}^{kl}(x) + \hat{p_t}^{kl}(x)\big(g_t(x) - \mathbb{E}_{\hat{p_t}^{kl}(x)}[g_t(x)]\big), \quad (23)$$

$$\begin{aligned}g_t(x) &= a\Delta f^{kl}(x) + a\langle\nabla\log p_t(x), \nabla f^{kl}(x)\rangle - \beta_t\frac{\sigma_t^2}{2}\Delta f^{kl}(x) + \frac{\partial\beta_t}{\partial t}f^{kl}(x) \\ &\quad + \left\langle \beta_t\nabla f^{kl}(x), v_t(x) - \sigma_t^2\nabla\log q_t(x) - \frac{\sigma_t^2}{2}\beta_t\nabla f^{kl}(x)\right\rangle. \quad (24)\end{aligned}$$

Taking $v_t(x) = -f_t(x) + \sigma_t^2\nabla\log q_t(x)$ and $a = \beta_t\sigma_t^2/2$, we have:

$$\begin{aligned}\frac{\partial \hat{p_t}^{kl}(x)}{\partial t} &= -\left\langle\nabla, \hat{p_t}^{kl}(x)\left(-f_t(x) + \sigma_t^2\nabla\log q_t(x) + \beta_t\frac{\sigma_t^2}{2}\nabla f^{kl}(x)\right)\right\rangle \\ &\quad + \frac{\sigma_t^2}{2}\Delta\hat{p_t}^{kl}(x) + \hat{p_t}^{kl}(x)\big(g_t(x) - \mathbb{E}_{\hat{p_t}^{kl}(x)}[g_t(x)]\big),\end{aligned}$$

$$\begin{aligned}g_t(x) &= \left\langle \beta_t\nabla f^{kl}(x), \frac{\sigma_t^2}{2}\nabla\log \hat{p_t}^{kl}(x)\right\rangle + \frac{\partial\beta_t}{\partial t}f^{kl}(x) + \left\langle \beta_t\nabla f^{kl}(x), -f_t(x) - \frac{\sigma_t^2}{2}\beta_t\nabla f^{kl}(x)\right\rangle \\ &= \frac{\partial\beta_t}{\partial t}f^{kl}(x) + \left\langle \beta_t\nabla f^{kl}(x), \frac{\sigma_t^2}{2}\nabla\log q_t(x) - f_t(x)\right\rangle. \quad (25)\end{aligned}$$

This can be simulated as:

$$dx_t = \left(-f_t(x_t) + \sigma_t^2\nabla\log q_t(x_t) + \beta_t\frac{\sigma_t^2}{2}\nabla f^{kl}(x_t)\right)dt + \sigma_t dW_t, \quad (26)$$

$$dw_t = \left[\frac{\partial\beta_t}{\partial t}f^{kl}(x_t) + \left\langle \beta_t\nabla f^{kl}(x_t), \frac{\sigma_t^2}{2}\nabla\log q_t(x_t) - f_t(x_t)\right\rangle\right]dt. \quad (27)$$

Step 2: Assuming $\beta_t = 1$, the above weighted SDE can be written as:

$$dx_t = \left(-f_t(x_t) + \sigma_t^2 \nabla \log q_t(x_t) + \frac{\sigma_t^2}{2} \nabla f^{kl}(x_t)\right) dt + \sigma_t dW_t, \tag{28}$$

$$dw_t = \left[\left\langle \nabla f^{kl}(x_t), \frac{\sigma_t^2}{2} \nabla \log q_t(x_t) - f_t(x_t) \right\rangle\right] dt. \tag{29}$$

By assuming $b_t(x_t) = -f_t(x_t) + \frac{\sigma_t^2}{2} \nabla \log q_t(x_t)$, we can write the following as,

$$dx_t = \left(b_t(x_t) + \frac{\sigma_t^2}{2} \nabla \log q_t(x_t) + \frac{\sigma_t^2}{2} \nabla f^{kl}(x_t)\right) dt + \sigma_t dW_t, \tag{30}$$

$$dw_t = \left[\left\langle \nabla f^{kl}(x_t), b_t(x_t) \right\rangle\right] dt. \tag{31}$$

Now, by following the standard SDE to ODE conversion rule, we can write as follows:

$$dx_t = \left(b_t(x_t) + \frac{\sigma_t^2}{2} \nabla f^{kl}(x_t)\right) dt, \tag{32}$$

$$dw_t = \left[\left\langle \nabla f^{kl}(x_t), b_t(x_t) \right\rangle\right] dt. \tag{33}$$

Now, utilizing the results from Proposition 4.2 regarding $-\nabla D_{KL}(q_t(x)||p_{boltz}(x))$ and utilizing the formulation of $f^{kl}(x_t) = -D_{KL}(q_t(x)||p_{boltz}(x))$, we get the final expression for the i-th particle as:

$$dx_t^i = b_t(x_t^i)\, dt + \frac{\sigma_t^2}{2} \frac{1}{n} \sum_{j=1}^{n} \left[k(x_t^j, x_t^i) \nabla_{x_t^j} V(x_t^j)\, dt + \nabla_{x_t^j} k(x_t^j, x_t^i)\, dt\right] \tag{34}$$

$$dw_t^i = \frac{1}{n} \sum_{j=1}^{n} \left[\langle k(x_t^j, x_t^i) \nabla_{x_t^j} V(x_t^j), b_t(x_t^i) \rangle\, dt + \langle \nabla_{x_t^j} k(x_t^j, x_t^i), b_t(x_t^i) \rangle\, dt\right] \tag{35}$$

This completes the Proof.

□

C Performance Comparison in Longer Horizon Search Tasks

While in the main paper, we evaluate IMPFM on feedback-efficient search tasks with relatively short horizons, this section extends our comparative analysis to longer-horizon tasks, with findings reported in Figure 10. Notably, the performance trends in these extended settings mirror those observed in the shorter-horizon tasks. These additional results reinforce that IMPFM is highly sample-efficient—achieving higher rewards with fewer interactions—while simultaneously maintaining a robust search capability over longer horizons that consistently outperforms the baselines.

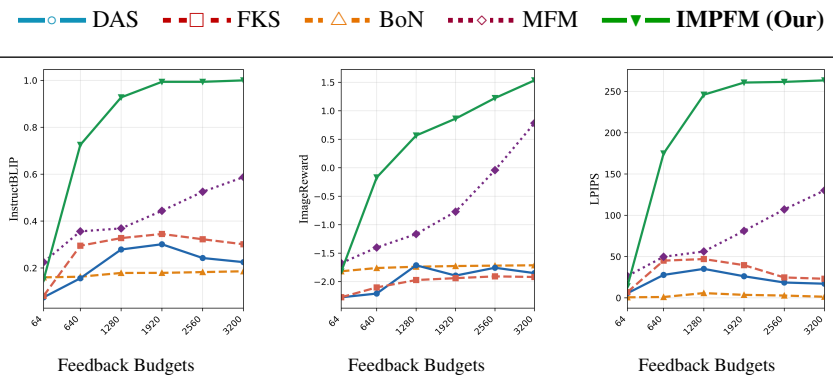


Figure 10: Search Performance Analysis of Competitive Approaches on Imagenet Target Classes.

D Additional Analysis on the Impact of the Interaction-Aware Feynman-Kac Corrector

Building on the qualitative and quantitative analysis presented in the main paper, we provide here additional results to further substantiate the impact of the interaction-aware Feynman-Kac corrector. Following the same ablation protocol, we remove the interaction terms — comprising both the repulsive term and the kernel-weighted value gradient — from the drift correction and reweighting mechanisms, yielding a standard reward-tilted Feynman-Kac corrector, referred to as FKC-Reward-Tilted. We benchmark this against our full Interaction-Aware Feynman-Kac corrector and report the results in Fig. 11. As illustrated, the interaction-aware formulation substantially enhances sample diversity, directly alleviating mode collapse. The key limitation of the FKC-Reward-Tilted sampler lies in its purely pointwise optimization objective: by neglecting global distributional constraints, it greedily pursues high-reward regions and consequently becomes trapped in local neighborhoods, rendering it susceptible to reward over-optimization. In contrast, the interaction-aware Feynman-Kac corrector incorporates a KL-divergence tilting factor that enforces distributional alignment, while the explicit inter-particle interaction mechanism actively promotes global coverage by aggregating information across the entire particle ensemble. This dual mechanism naturally steers particles away from local optima, fostering both diversity and robustness against reward over-optimization, as clearly reflected in Fig. 11. Furthermore, we analyze the specific impact of the interaction term within the drift correction mechanism (Eq. 6). To isolate this effect, we visually compare the samples generated by our fully interactive formulation (MPC) against an ablated, non-interacting baseline (MFM). As shown in Figure 12, incorporating particle interaction significantly improves both sample diversity and target alignment. This highlights the interaction term as a critical component for overcoming the severe mode collapse and reward over-optimization typical of standard optimal control frameworks. In a nutshell, these additional qualitative analyses reinforce the importance of the interaction-aware Feynman-Kac corrector for feedback-efficient search.



Figure 11: Impact of Particle Interaction in Mitigating Reward Over-optimization on FKC.

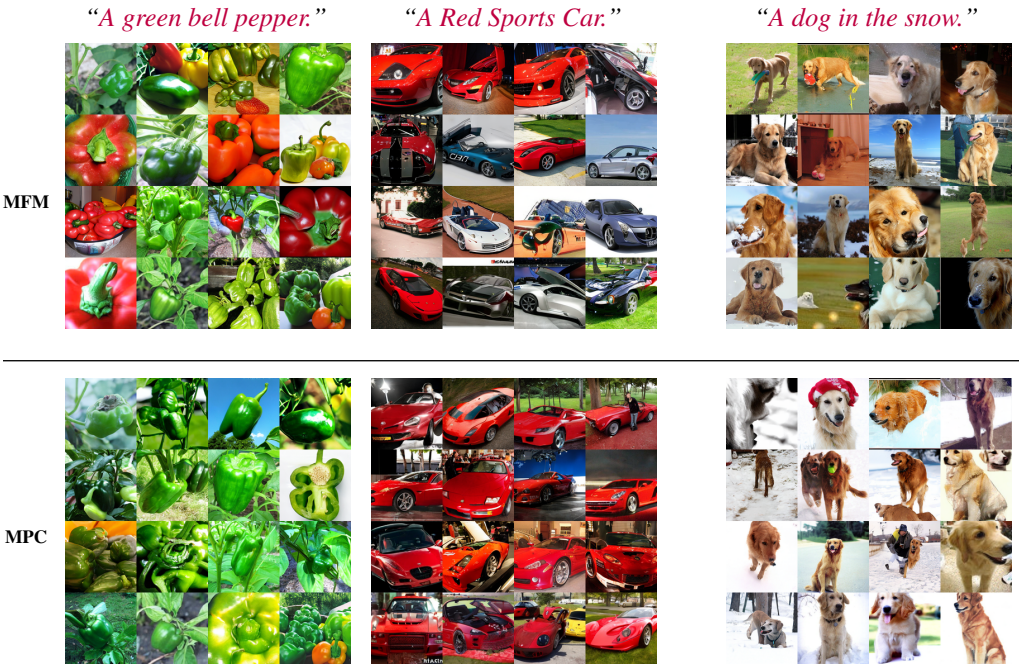


Figure 12: Impact of Particle Interaction in Mitigating Reward Over-optimization on Optimal Control.

E VISUALIZING THE IMPACT OF SUFFICIENT STATISTIC

In the main paper (Section 4), we analyze the impact of the Sufficient Statistic (SS). Here, we present additional visualizations highlighting that the SS is crucial for enabling feedback-efficient search, especially under severe budget constraints. As depicted in Fig. 13, when the feedback step is less than 50, the alternative approach fails to generate high-quality samples. In contrast, *by leveraging dynamic transition steps enabled via the SS mechanism, IMPFM successfully generates high-quality, target-aligned samples even with a very small feedback step (e.g., 20).*

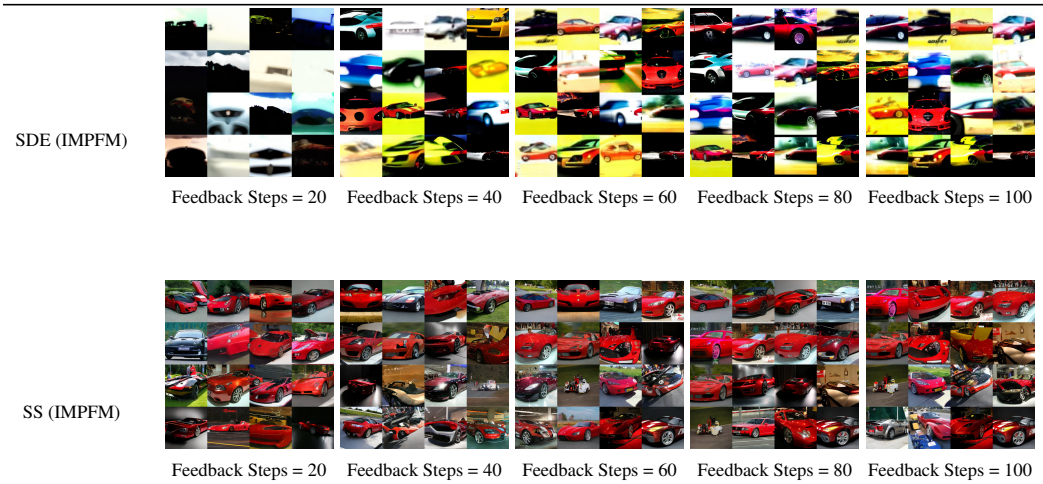


Figure 13: Visualizing the Impact of Sufficient Statistics for Lower Feedback Budget.

F Additional Qualitative Analysis: Taming Weight Degeneracy Through Particle Interaction-Guided Drift Correction

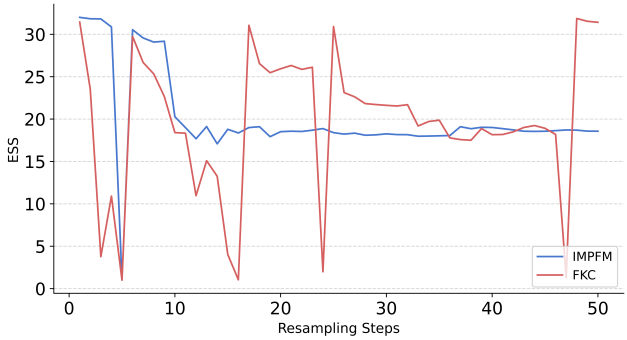


Figure 14: Efficacy of Interactive Particle mechanism on Mitigating Weight Degeneracy on FKC Sampler.



Figure 15: Impact of Interactive Multi-Particle Drift Correction Mechanism on Mitigating Weight Degeneracy.

In this section, we present a deeper qualitative investigation into how the proposed multi-particle interaction-driven drift correction mechanism combats weight degeneracy. To achieve this, we visualize the denoised samples generated via flow-map for each particle in the batch and compute the Effective Sample Size (ESS) at every single transition step. By contrasting our fully interacting dynamics (Eq. 6) against a naive baseline—where particles are treated in strict isolation and corrected purely based on individual value—we directly highlight the stabilizing power of our approach. We report the findings in Figure 14 and 15. Crucially, we demonstrate that multi-particle interaction acts as a fundamental driver of sample diversity. Unlike isolated particle dynamics—which rapidly collapse into semantically redundant outputs—our interacting system sustains rich, diverse generation across every correction step. This structural stability is corroborated by a drastic, consistent improvement in the Effective Sample Size (ESS), effectively eradicating the pervasive weight degeneracy problem. Importantly, this interactive framework is made computationally viable by flow-map; by enabling

highly efficient evaluation of value gradients (∇V_t), flow-maps unlock the ability to compute complex, multi-particle interactions without prohibitive overhead.

G Details of Class and Prompt Detail in the Search Experiment

In the following table 1, we detail the randomly selected ImageNet target classes and their fine-grained specifications used to evaluate the competing methods on the feedback-efficient search task.

Target Class	**Fine-Grained Prompt**
Car	A Red Sports Car
Cat	A Black Cat.
Dog	A Dog in the Snow.
Daisy	A Red Daisy.
Pepper	A Green Bell Pepper.

Table 1: Target classes and their corresponding fine-grained prompts for the Feedback-efficient Search task.

H Additional Comparative Visualization on Search Task

In Fig. 16, we present additional comparative visualizations for search tasks involving diverse classes from ImageNet. These figures further highlight the impact of IMPFM in enabling feedback-efficient search.

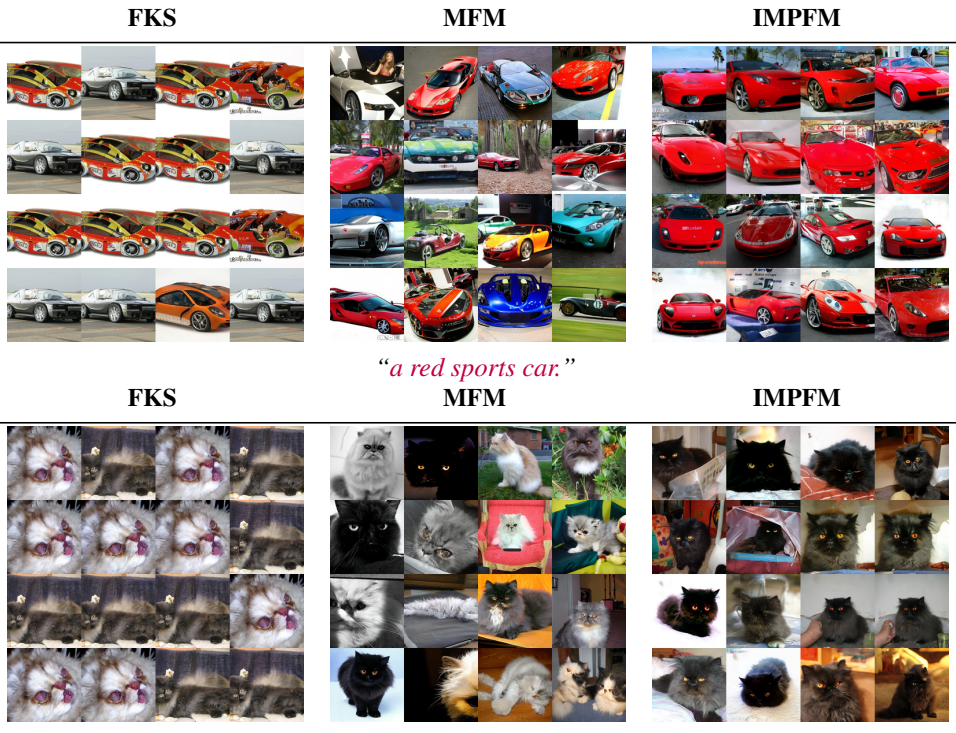


Figure 16: Search Visualizations.

I Details about the Evaluation Dataset for Alignment Tasks

We provide the evaluation prompts for both the compositional and quantity-aware alignment tasks in separate Excel files, available via the anonymous link.

J Implementation Details

For all experiments in the feedback-driven search setting, we use 32 particles, with each particle contributing one posterior sample for feedback guidance, resulting in 32 feedback samples per resampling step. Notably, when configured with a single particle, IMPFM reduces to the MFM baseline. However, as the particle count increases, IMPFM outperforms MFM under the same feedback budget due to its posterior sample sharing mechanism. Other implementation details are available in the provided code. Finally, all experiments are implemented in PyTorch and conducted on 3 NVIDIA A100 GPUs. Our training and inference code will be made public.

K Derivation of TrigFlow/SANA (Chen et al., 2025) Noise Schedule to Adapt IMPFM on Alignment Tasks

In this section, we describe how we adapt the trigonometric noise schedule from TrigFlow (Chen et al., 2025) and the consistency model (SANA-Sprint) (Chen et al., 2025) to our IMPFM framework, which operates under a different noise schedule. We first review the TrigFlow parameterization, then describe the consistency model reparameterization, and finally derive the adaptation of the noise schedule to our setting.

K.1 Background: TrigFlow Transformation

TrigFlow introduces a trigonometric reparameterization of the Flow Matching timestep, establishing a bijective mapping between the standard FM timestep $t_{FM} \in [0, 1]$ and a cosine-based schedule parameterized by t_{cm}. Specifically, the mapping is defined as:

$$t_{FM} = \frac{\sin(t_{cm})}{\sin(t_{cm}) + \cos(t_{cm})}, \tag{36}$$

where t_{cm} controls the balance between noise and signal through trigonometric interpolation. This mapping ensures that $t_{FM} = 0$ corresponds to a clean sample and $t_{FM} = 1$ corresponds to pure noise, while providing a smoother, geometry-aware interpolation path compared to a linear schedule.

Given the mapping in Eq. equation 36, the noisy latent $X_{t_{FM}}$ is expressed in terms of the original latent $X_{t_{cm}}$ as:

$$X_{t_{FM}} = \frac{X_{t_{cm}}}{\sigma_d} \sqrt{t_{FM}^2 + (1 - t_{FM})^2}, \tag{37}$$

where σ_d is a scaling factor that normalizes the latent representation. The factor $\sqrt{t_{FM}^2 + (1 - t_{FM})^2}$ accounts for the norm of the trigonometric interpolation coefficients, ensuring consistent scaling across timesteps.

Under this reparameterization, the score network $\hat{F}_\theta$ is reformulated to operate on the rescaled latent $X_{t_{cm}}/\sigma_d$ at timestep t_{cm}. The relationship between the TrigFlow score network and the velocity field V_θ of the standard Flow Matching model is:

$$\begin{aligned} \hat{F}_\theta\left(\frac{X_{t_{cm}}}{\sigma_d},\, t_{cm},\, y\right) = {} & \frac{1}{\sqrt{t_{FM}^2 + (1 - t_{FM})^2}} \\ & \times \Big[(1 - 2t_{FM})\, X_{t,FM} + \left(1 - 2t_{FM} + 2t_{FM}^2\right) V_\theta(X_{t_{FM}},\, t_{FM},\, y)\Big], \end{aligned} \tag{38}$$

where y is the conditioning input (e.g., a text prompt). This reformulation shows that the TrigFlow score network can be expressed as a linear combination of the noisy latent and the predicted velocity field, weighted by time-dependent coefficients that are derived analytically from the trigonometric schedule.

K.2 SANA-SPRINT REPARAMETERIZATION

SANA-Sprint builds upon TrigFlow by introducing a consistency-style denoising function f_θ that directly maps a noisy latent X_t to a clean estimate $\hat{X}_0$ in a single step. Under the TrigFlow time parameterization, this function takes the form:

$$f_\theta(X_t, t) = \cos(t_{cm}) \cdot X_t - \sin(t_{cm}) \cdot \sigma_d \cdot F_\theta\left(\frac{X_t}{\sigma_d},\, t_{cm}\right), \tag{39}$$

where $X_{cm} = X_t/\text{Scale}$ is the rescaled latent. Intuitively, Eq. equation 39 can be understood as a trigonometric decomposition: the $\cos(t_{cm})$ term retains the signal component of X_t, while the $\sin(t_{cm})$ term removes the noise component as predicted by F_θ. This yields the denoised estimate:

$$\hat{X}_0 = \cos(t_{cm})\, X_{t_{cm}} - \sin(t_{cm})\, \hat{F}_\theta, \tag{40}$$

which provides a clean, geometry-consistent reconstruction of the original sample from any noisy intermediate state.

K.3 ADAPTING TO OUR IMPFM FRAMEWORK WITH THE SUFFICIENT STATISTIC (S) SCHEDULE

Our IMPFM framework uses a different noise schedule compared to TrigFlow. Specifically, we adopt the S schedule, which defines a signal-to-noise ratio (SNR) function that differs from the trigonometric parameterization used in TrigFlow. To adapt the SANA-Sprint framework to our setting, we must establish a consistent mapping between the two schedules.

Setting up the Schedule Correspondence. Under the S schedule, the signal and noise coefficients at timestep $t = t_{cm}$ are defined as:

$$\alpha_t = \cos t = \sqrt{s}, \qquad \sigma_t = \sin t = \sqrt{1-s}, \tag{41}$$

where $s \in [0, 1]$ is the schedule variable. This gives the forward noising process:

$$X_t = \left(\alpha_t\, \hat{X}_0 + \sigma_t\, \frac{\varepsilon}{\sigma_d}\right) \cdot \sigma_d = \cos(t)\, \hat{X}_0 \cdot \sigma_d + \sin(t)\, \varepsilon, \tag{42}$$

which is a standard affine interpolation between the clean sample $\hat{X}_0$ and Gaussian noise $\varepsilon \sim \mathcal{N}(0, I)$, scaled by the latent normalization factor σ_d.

Defining the SNR Function. Under the S schedule, the SNR function $g(s)$ and its transformed counterpart $\tilde{g}(\alpha_t)$ are defined as:

$$g(s) = \frac{1}{s} - 1, \qquad \tilde{g}(\alpha_t) = \frac{1}{\alpha_t} - 1 = \frac{1}{\cos^2(t_{cm})} - 1, \tag{43}$$

where $\tilde{g}(\alpha_t)$ is the SNR expressed as a function of the signal coefficient $\alpha_t = \cos(t_{cm})$. This function is monotonically decreasing in t_{cm}, consistent with the intuition that noise increases as the timestep progresses.

Deriving the Optimal Transition Timestep $r^* = r^*_{cm}$ for computing DDPM transition via Sufficient Statistic. A key step in adapting the noise schedule is determining the optimal intermediate timestep r^* at which to perform the consistency transition between two noise levels t_{cm} and $t_{cm'}$ (with $t_{cm'} < t_{cm}$, i.e., $t_{cm'}$ is closer to the clean data). The optimal r^* is derived by matching the SNR levels under the S schedule, yielding:

$$\tilde{g}(r^*) = \frac{\tilde{g}(\cos t_{cm'}) \cdot \tilde{g}(\cos t_{cm})}{\tilde{g}(\cos t_{cm}) - \tilde{g}(\cos t_{cm'})}, \tag{44}$$

which can be solved by applying the inverse of g to both sides:

$$\hat{r}^* = g^{-1}\left(\frac{\tilde{g}(\cos t_{cm'}) \cdot \tilde{g}(\cos t_{cm})}{\tilde{g}(\cos t_{cm}) - \tilde{g}(\cos t_{cm'})}\right). \tag{45}$$

Eq. equation 45 gives the optimal SNR level r^* in the s-domain. To recover the corresponding trigonometric timestep r^*_{cm}, we use the relation $\alpha_t = \cos(t_{cm}) = \sqrt{s}$, which gives:

$$\cos(r^*_{cm}) = d\sqrt{r^*} \quad \Longrightarrow \quad r^*_{cm} = \arccos\left(\sqrt{r^*}\right). \tag{46}$$

This recovers the optimal intermediate timestep r^*_{cm} in the trigonometric domain, which can be directly used within the TrigFlow/SANA-Sprint framework.

Computing the DDPM Transition via Sufficient Statistic Given the optimal transition timestep r^*_{cm}, the DDPM transition sample $X_{t_{cm'}}$ at the new timestep $t_{cm'}$ is obtained by applying the Sufficient Statistic scheme as defined in Equation 9 in the main paper.

L Sufficient Statistic Computation

Sufficient Statistic computation requires to evaluate $\bar{\gamma}$, ρ and σ. Following (Holderrieth et al., 2025), it can be computed as $\bar{\gamma} = \rho\sigma_{t^{child}}\sigma_t^{-1}$, $\bar{\alpha} = \alpha_{t^{child}} - \bar{\gamma}\alpha_t$, $\bar{\sigma}^2 = \sigma^2_{t^{child}}(1-\rho^2)$. ρ is the correlation parameter (Holderrieth et al., 2025).

M Extended Related Work

Fine-Tune Generative model with RL: Fine-tuning generative models with human feedback, such as user preferences, has become increasingly prevalent (Ouyang et al., 2022; Touvron et al., 2023). However, these approaches typically assume a static reward model, treating rewards as fixed ground truth and not accommodating online queries. In contrast, we study an online, interactive setting in which target properties are initially unknown and are gradually revealed through sequential feedback, enabling continual target discovery rather than one-shot alignment to a fixed reward proxy. (Dong et al., 2023) introduces a general online learning framework for aligning generative models, but it is not tailored to diffusion models. More recently, (Uehara et al., 2024) proposed a feedback-efficient online fine-tuning method for diffusion models; however, it learns a separate online reward model to guide sampling, which amplifies bias in the early stages of fine-tuning, and due to the KL-regularized fine-tuning objective, it remains vulnerable to mode collapse. While tree-based methods Jain et al. (2025) provide a gradient-free alternative for search, they lack a mechanism for active trajectory correction. By relying strictly on prior dynamics, these samplers fail to leverage acquired feedback effectively, resulting in inefficient search under limited budget constraints.

N Background: Feynman-Kac Correctors

Feynman-Kac Corrector (FKC) framework has been proposed to effectively sample from modified target distributions at inference time—such as reward-tilted distributions, annealed densities, or geometric averages (Skreta et al., 2025). In standard diffusion models (Song et al., 2020b;a), the time-evolution of the sample density $p_t(x)$ under the reverse-time stochastic differential equation (SDE) is governed by the Fokker-Planck partial differential equation (PDE):

$$\frac{\partial p_t(x)}{\partial t} = -\langle \nabla, p_t(x) v_t(x) \rangle + \frac{\sigma_t^2}{2} \Delta p_t(x) \tag{47}$$

where $v_t(x) = -f_t(x) + \sigma_t^2 \nabla \log q_t(x)$ represents the reverse-time drift. However, heuristically modifying this drift to target new compositional distributions fails to track the prescribed intermediate marginals.

The FKC method corrects this by formulating the exact time-evolution of the target distribution as a Feynman-Kac PDE. This extends the standard Fokker-Planck equation by introducing a spatial reweighting function, $\bar{g}_t(x) = g_t(x) - \int g_t(x) p_t(x) dx$, which rigorously quantifies the mismatch between the simulated heuristic and the true target marginals. The combined evolution is described by:

$$\frac{\partial p_t^{\text{FK}}(x)}{\partial t} = -\langle \nabla, p_t^{\text{FK}}(x) v_t(x) \rangle + \frac{\sigma_t^2}{2} \Delta p_t^{\text{FK}}(x) + \bar{g}_t(x) p_t^{\text{FK}}(x) \tag{48}$$

Consequently, the sampling process is cast as a weighted SDE, where a particle x_t and its accumulated log-weight w_t evolve according to:

$$dx_t = v_t(x_t)dt + \sigma_t dW_t, \quad dw_t = \bar{g}_t(x_t)dt \tag{49}$$

Crucially, this mathematical framework seamlessly extends to reward-tilted sampling, where the objective is to generate samples from $p_t^{\text{reward}}(x) \propto q_t(x)\exp(\beta_t r(x))$ for a differentiable reward function $r(x)$ and inverse temperature β_t. Under the FKC formulation, the drift is augmented by the reward gradient to proactively steer the simulation, and the reweighting function accumulates mass based on the alignment between the diffusion vector field and the reward gradient. The resulting weighted SDE for reward-tilted sampling is exactly specified as:

$$dx_t = \left(-f_t(x_t) + \sigma_t^2 \nabla \log q_t(x_t) + \beta_t \frac{\sigma_t^2}{2}\nabla r(x_t)\right) dt + \sigma_t dW_t \tag{50}$$

$$dw_t = \left[\frac{\partial \beta_t}{\partial t} r(x_t) + \left\langle \beta_t \nabla r(x_t), \frac{\sigma_t^2}{2}\nabla \log q_t(x_t) - f_t(x_t)\right\rangle\right] dt \tag{51}$$

To efficiently simulate these weighted equations and prevent weight degeneracy, the FKC framework employs Sequential Monte Carlo (SMC) resampling. By simulating a batch of non-interacting particles and periodically resampling them proportional to their exponentiated weights $\exp(w_t)$, the population dynamically corrects its distribution to faithfully match $p_t^{\text{reward}}(x)$ throughout the reverse trajectory.

O Additional Visualizations on the Alignment Task

In Fig. 18, 19, 20, we provide additional comparative visualizations for the quantitative alignment tasks. These results underscore the effectiveness of IMPFM in complex, feedback-efficient alignment tasks that demand robust reasoning capabilities.



Figure 17: Alignment Visualization. (left) Quantity Prompt. (right) Compositional Prompt.

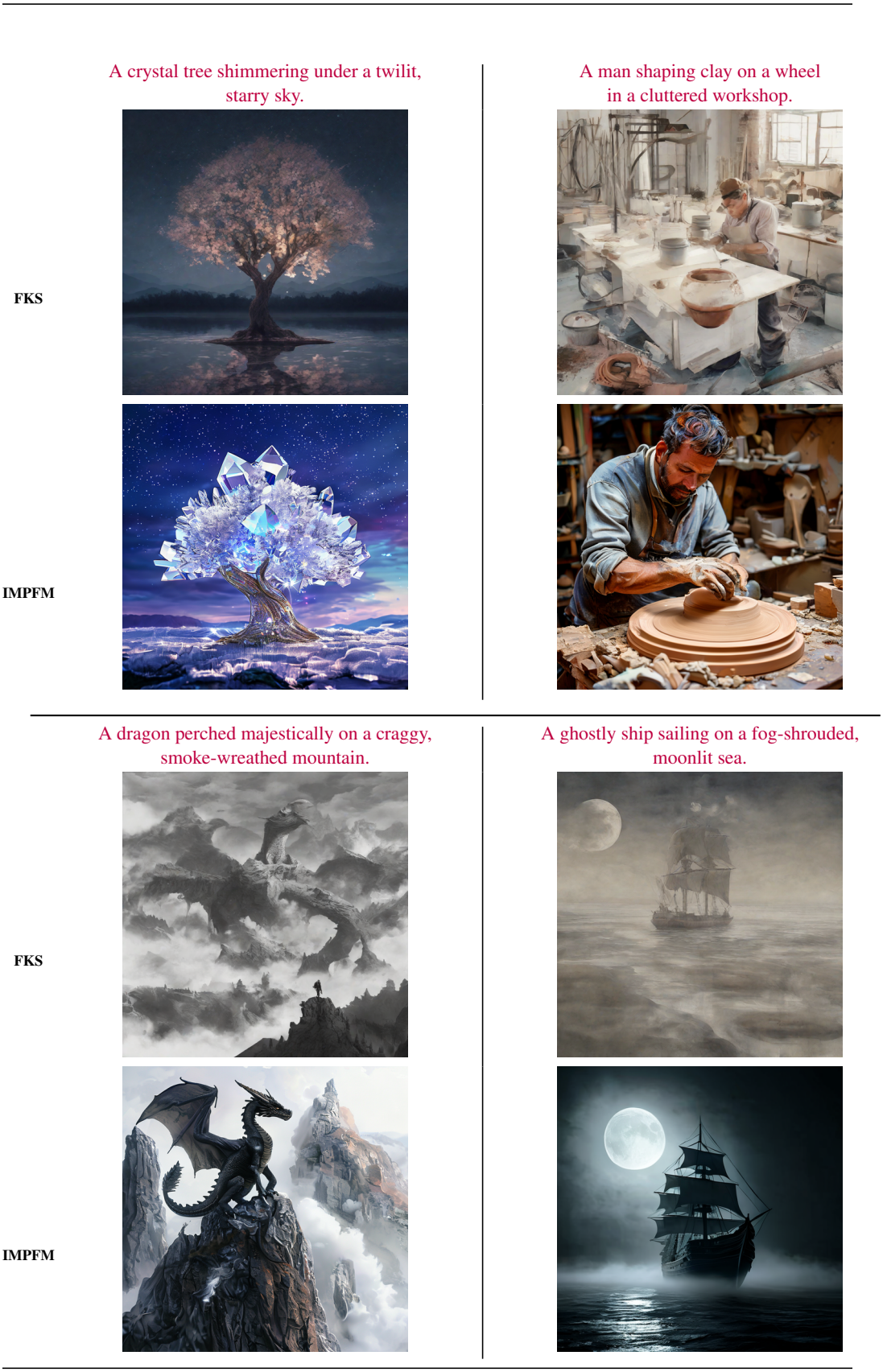


Figure 18: Additional Alignment Visualization.

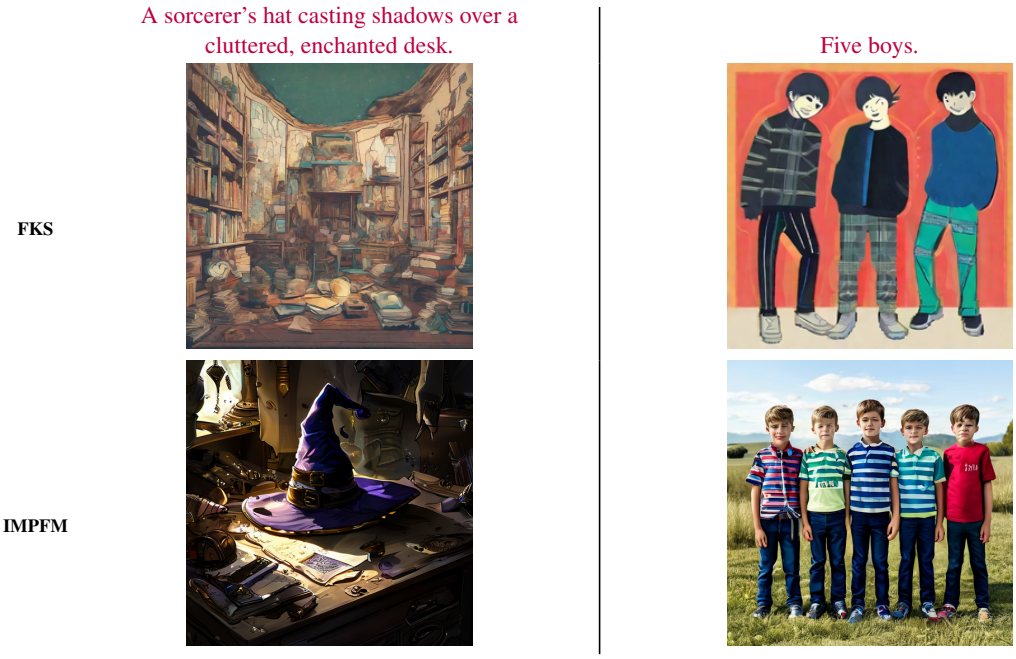


Figure 19: Additional Alignment Visualization.



Figure 20: Additional Alignment Visualization.