Active Flow Expansion for Out-of-Distribution Discovery: from Theory to Molecules

Riccardo De Santi [1 2] **Bruce Lee** [1 2] **Cristian Perez Jensen** [1] **Kimon Protopapas** [1] **Sophia Tang** [3]
Cheng-Hao Liu [4 5] **Pranam Chatterjee** [3] **Yisong Yue** [4] **Andreas Krause** [1 2]

Abstract

Standard flow and diffusion pre-training matches the distribution of available data (e.g., molecules), which often covers only a small fraction of the valid design space. In generative discovery, however, one aims to sample valid new-to-nature designs, assigned negligible probability under, and thus inaccessible to, standard models fitted to the observed data. To overcome this limitation, we depart from data distribution matching and view a generative model through its *generable set*: the region it covers with non-negligible probability. This allows to introduce a new learning principle for *out-of-distribution flow modeling*: enlarging a model's generable set to increase coverage of the valid design space. We propose *Active Flow Expansion (*ActFlow*)*, a continued pre-training method that employs verifier feedback to expand a pre-trained model over new valid regions by iteratively adapting to synthetic data generated through active exploration in the learned flow representation. Theoretically, we establish to our knowledge first-of-their-kind statistical learning guarantees for out-of-distribution flow modeling, analyzing generable set expansion as a local-to-global reachability process over a learned representation. Empirically, we assess ActFlow with suitable out-of-distribution generative modeling metrics across small organic molecules, mid-sized drug-like molecules, therapeutic peptides, and protein sequence design tasks. Results show that ActFlow expands valid coverage far beyond the region modeled by the initial pre-trained model, significantly outperforming widely adopted synthetic flow pre-training methods.

[1]ETH Zurich [2]ETH AI Center [3]University of Pennsylvania [4]Caltech [5]FutureHouse. Correspondence to: Riccardo De Santi <rdesanti@ethz.ch>.

Accepted at the 2026 Workshop on Generative and Agentic AI for Biology (ICML 2026)

1. Introduction

Large-scale generative modeling has advanced rapidly in recent years, with flow (Lipman et al., 2022; 2024) and diffusion models (Sohl-Dickstein et al., 2015; Song & Ermon, 2019; Ho et al., 2020) emerging as powerful approaches for generating high-fidelity samples across domains including chemistry (Hoogeboom et al., 2022), biology (Corso et al., 2022), and robotics (Chi et al., 2025). Alongside this progress, over the last years, a growing literature has introduced test-time reward-tilting techniques (e.g., Uehara et al., 2025; Perez Jensen et al., 2026; Uehara et al., 2024; De Santi et al., 2025b). These advances have expanded the use of generative models in scientific discovery, enabling applications such as molecular property optimization (Gutjahr et al., 2025), or functional enzyme design (Rector-Brooks et al., 2026).

A Core Limitation of Standard Flow Pre-training for Discovery. Despite recent progress, generative discovery remains largely constrained by standard pre-training. Flow and diffusion models are learned by matching the distribution of available data (Song et al., 2020; Lipman et al., 2024), which typically covers only a limited portion of the valid design space. This objective is natural when the goal is to reproduce observed examples, e.g., in robotics for manufacturing, where behavioral cloning is often sufficient (Pearce et al., 2023), but it is poorly suited to out-of-distribution discovery, where one typically seeks valid new-to-nature designs with negligible probability under the data distribution, and arbitrarily far from available data acquired via nature's evolution and prior discoveries. In other words, the goal of out-of-distribution discovery is fundamentally opposed to standard generative modeling as distribution matching. Nonetheless, pre-trained generative models remain, to date, arguably the most promising way to access the complex design spaces in which scientific discovery takes place. This raises the central question of this work:

How can we adapt a pre-trained flow or diffusion model in a task-agnostic way to enable out-of-distribution discovery?

Answering this question would advance the algorithmic-theoretical foundations of generative discovery, with

	GEOM-Drugs Molecules		Therapeutic Peptides		Protein Sequences	
Method	**Coverage ↑**	**Diversity ↑**	**Coverage ↑**	**Diversity ↑**	**Coverage ↑**	**Diversity ↑**
Pre-trained	35.89 ± 2.05	255.03 ± 1.09	133.30 ± 11.1	5.70 ± 0.11	66.50 ± 5.63	12.87 ± 0.63
REC-NF	44.67 ± 2.36	267.10 ± 2.88	0.00 ± 0.00	0.00 ± 0.00	63.75 ± 11.45	11.85 ± 0.34
REC-F	89.33 ± 11.56	284.30 ± 2.62	45.00 ± 37.22	6.03 ± 1.19	49.50 ± 9.60	11.67 ± 0.40
ACTFLOW	**144.30** ± 19.28	**303.10** ± 5.71	**197.00** ± 66.25	**7.45** ± 6.17	**102.75** ± 18.36	**42.14** ± 10.85

Table 1. ACTFLOW expands model coverage (i.e., valid clusters) and diversity (i.e., Vendi (Friedman & Dieng, 2022) of valid samples) across domains from de novo molecular design to protein sequence design, significantly outperforming widely adopted baselines.

direct implications for high-impact scientific discovery applications. We take a step toward this goal by moving beyond data distribution matching and *rethinking flow model learning* for out-of-distribution generative modeling. Concretely, we make the following contributions.

Our contributions.

- We introduce the notion of *generable set*: the region of design space a model can access with non-negligible probability. This yields a rigorous mathematical framework for *out-of-distribution flow modeling*, and a new learning principle for continued pre-training: expanding a model generable set to cover a larger portion of the valid design space, which we name *generable set expansion*.
- We propose Active Flow Expansion (ACTFLOW), a continued pre-training method that expands a pre-trained flow model through recurrent adaptation on self-generated synthetic data – as illustrated in Fig. 1 (right). ACTFLOW actively explores in the *learned diffusion representation* at intermediate noise levels, where the geometry is more favorable to global exploration.
- We establish, to our knowledge, first-of-their-kind statistical learning guarantees for out-of-distribution flow modeling. These rely on an energy-based modeling abstraction, and analyze active generable set expansion as a local-to-global reachability process in a learned representation.
- We employ domain-agnostic metrics for out-of-distribution generative modeling and assess ACTFLOW across small organic molecules, mid-sized drug-like molecules, therapeutic peptides, and protein design. ACTFLOW significantly increases valid coverage beyond the initial pre-trained model, vastly outperforming widely used recursive data generation baselines. We further show that the same algorithmic principles and empirical gains extend to discrete (diffusion) models.

2. From Data Distribution Matching to Out-of-Distribution Flow Modeling

We consider a design space $\mathcal{X} \subseteq \mathbb{R}^d$, where each $x \in \mathcal{X}$ is a candidate design, such as a molecule or protein. Let p_{data} denote the distribution of the available data, assumed to consist of i.i.d. samples that satisfy the domain-specific notion of validity, e.g., valid molecules. A continuous-time flow model is defined by a time-dependent velocity field $u_\theta : \mathcal{X} \times [0,1] \to \mathcal{X}$ and the ordinary differential equation

$$\frac{dx_t}{dt} = u_\theta(x_t, t), \qquad x_0 \sim p_0, \tag{1}$$

where p_0 is a simple base distribution, typically Gaussian (Lipman et al., 2024). Let p_t^θ denote the marginal density induced at time t by the flow, and let p_1^θ denote its terminal density. In standard flow matching pre-training, one specifies conditional interpolation paths $p_t(\cdot \mid x_0, x_1)$ between $x_0 \sim p_0$ and $x_1 \sim p_{data}$, with associated target velocity field $u_t^\star(\cdot \mid x_0, x_1)$, and learns u_θ by minimizing the loss:

$$\mathbb{E}\left[\|u_\theta(x_t, t) - u_t^\star(x_t \mid x_0, x_1)\|_2^2\right], \tag{2}$$

$$t \sim U[0,1],\ x_t \sim p_t(\cdot \mid x_0, x_1). \tag{3}$$

This vector-field regression objective allows to match the terminal distribution p_1^θ to the available-data distribution p_{data}, as by Equation 4. Beyond flow matching, the same distribution-matching learning principle underlies also standard pre-training of continuous (Song et al., 2020) and discrete (Lou et al., 2023) diffusion models.

Distribution Matching: Standard Flow Model Pre-Training

$$\text{learn } \theta \text{ such that } p_1^\theta \approx p_{data}. \tag{4}$$

2.1. Generable Set Expansion: Flow Modeling Beyond Data Distribution Matching

In this work, we depart from viewing generative models as data distribution approximators and instead cast them as *valid design space approximators*. Let $v : \mathcal{X} \to \{0,1\}$ denote whether a design is valid according to the domain one wishes to model, e.g., v might express molecular physiochemical validity, protein sequence foldability, etc. We indicate via Ω^* the entire valid design space:

$$\Omega^* := \{x \in \mathcal{X} : v(x) = 1\}, \tag{5}$$

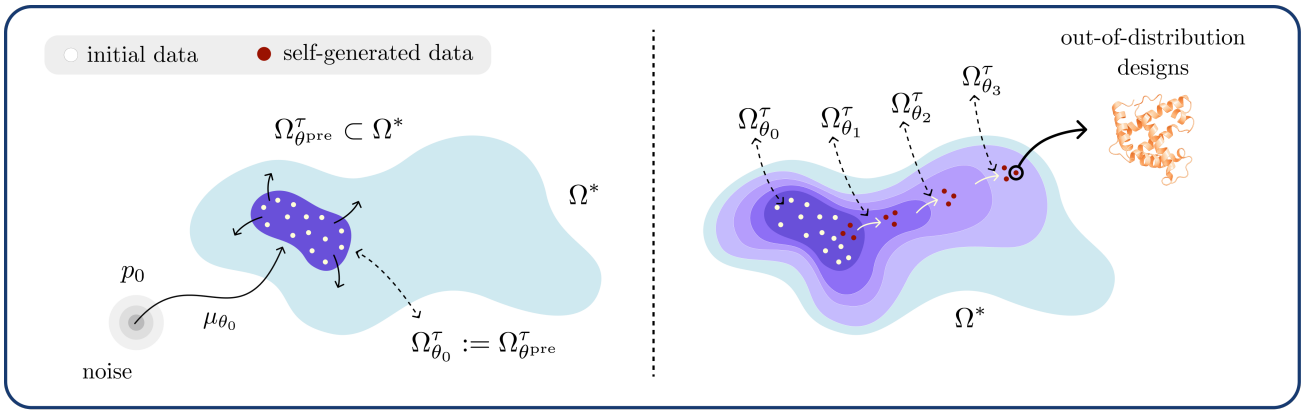


Figure 1. (left) Pre-trained flow model μ_{θ_0} generates with sufficient probability a set $\Omega^{\tau}_{\theta_0}$ poorly covering the valid design space Ω^*, i.e., $\Omega^{\tau}_{\theta_0} \subset \Omega^*$. (right) Active Flow Expansion (ACTFLOW) increases model coverage over new valid regions of Ω^*, enabling out-of-distribution flow modeling.

which we wish our generative model to cover well. To capture the data regions sufficiently covered by a pre-trained flow model μ_θ, we introduce the following notion of *generable set*.

Definition 1 (Generable Set)**.** *Let μ_θ be a flow model inducing terminal density p_1^θ on $\mathcal{X}$. For $\tau > 0$, we define its τ-level* generable set *as*

$$\Omega^{\tau}_{\theta} := \{x \in \mathcal{X} \mid p_1^\theta(x) \geq \tau\}. \tag{6}$$

For $\tau \to 0$ the generable set approximates the model support, i.e., $\Omega^0_\theta := \text{Supp}(p_1^\theta)$. However, the support can include regions assigned only infinitesimal, yet positive probability, and therefore may substantially overstate the model's effective coverage. Instead, for sufficiently small $\tau > 0$, the generable set Ω^{τ}_{π} captures the region likely to be sampled under a finite budget. With this notion of generable set, we can state the following limitation of standard pre-trained flows for out-of-distribution discovery [1]:

Central Limitation of Standard Flow Pre-training

$$\Omega^{\tau}_{\theta^{\text{pre}}} \subset \Omega^* \qquad \text{with} \qquad \text{Vol}(\Omega^{\tau}_{\theta^{\text{pre}}}) \ll \text{Vol}(\Omega^*). \tag{7}$$

Under exact distribution matching, Eq. (7) allows to define the valid out-of-distribution (OOD) region:

$$\text{Valid OOD region:} \quad \bar{\Omega}^{\tau}_{\theta^{\text{pre}}} := \Omega^* \setminus \Omega^{\tau}_{\theta^{\text{pre}}} \tag{8}$$

This is the set of valid designs not reliably covered by the pre-trained model $\mu_{\theta_{\text{pre}}}$ and typically not represented in p_{data}, such as new-to-nature designs. To overcome this limitation, we introduce *generable set expansion* (see Eq. 9), a new learning principle for continued pre-training: given a standard pre-trained flow model, expand its generable set to cover a larger portion of the valid design space Ω^*.

Generable Set Expansion: Out-of-Distribution Flow Modeling

$$\text{learn } \theta \text{ such that } \Omega^{\tau}_{\theta} \approx \Omega^*. \tag{9}$$

2.2. Generable Set Expansion via Iterative Continued Pre-training

Given a pre-trained flow model $\mu_{\theta^{\text{pre}}}$, we approach generable set expansion through continued pre-training on self-generated data. Beyond the data distribution, validity must be obtained through black-box verifier queries, e.g., molecular validity (De Santi et al., 2026; Guo & Schwaller, 2024) or protein foldability (Jumper et al., 2021; Watson et al., 2023). Since such queries can only be made on generable designs, expansion is intrinsically recursive, as shown in Fig. 1:

$$\theta^{\text{pre}} = \theta_0 \to \theta_1 \to \cdots \to \theta_T, \tag{10}$$

$$\Omega^{\tau}_{\theta_0} \subseteq \Omega^{\tau}_{\theta_1} \subseteq \cdots \subseteq \Omega^{\tau}_{\theta_T} \approx \Omega^\star. \tag{11}$$

The central algorithmic question is therefore how to self-generate data that expands the model's generable set, rather than merely reinforcing regions already assigned high density. Standard sampling is poorly suited to this goal: it preferentially draws from dominant modes, and may therefore neglect valid frontier regions that enable valid expansion. We report an illustrative Gaussian analysis in Apx. B.1 to isolate this failure mode in a minimal setting. The analysis shows that, even when expansion-enabling samples remain within the current generable set, passive self-generation can be exponentially unlikely to sample them when valid

[1] For clarity, here we consider a model generating only valid designs. In general $\Omega^{\tau}_{\theta^{\text{pre}}} \cap (\mathcal{X} \setminus \Omega^*)$ is non-empty (Apx. A)

directions of expansion are sparse. This highlights the core algorithmic requirement: self-generation must steer sampling toward generable and valid *frontier* regions, rather than merely reinforce dominant modes. In the next section, we introduce ACTFLOW, which implements this principle and yields provable generable set expansion guarantees.

3. Algorithm: Active Flow Expansion

We now introduce Active Flow Expansion (ACTFLOW) (Alg. 1), a continued pre-training method for *generable set expansion*: it expands a pre-trained flow model's generable set to cover a larger portion of the valid design space. Concretely, ACTFLOW fine-tunes the model on self-generated data acquired via inference-time active exploration in a learned flow representation at intermediate noising levels.

High-level Algorithm Summary. At round t, ACTFLOW fits a verifier uncertainty estimate $\sigma_t(\cdot)$ from a buffer $\mathcal{D}_t = \{(x_i, y_i)\}_{i=1}^t$, self-generates a new candidate, or a batch, x_{t+1} via Eq. (12), queries the verifier to obtain $y_{t+1} = \tilde{v}(x_{t+1})$, appends (x_{t+1}, y_{t+1}) to $\mathcal{D}_t$, and updates the flow into $\mu_{\theta_{t+1}}$.

Active Exploration over a Noised Flow Representation. ACTFLOW performs active self-generation in the learned representation $\mathcal{Z}_s$, given by hidden features of the velocity network μ_θ at noising level $s \in (0,1)$. At each iteration, Algorithm 1 steers the current model at inference-time toward high verifier-uncertainty regions in $\mathcal{Z}_s$, while remaining close to the current model density:

(Inference-Time) Active Exploration over Noised Flow Representation $\mathcal{Z}_s$

$$x_{t+1} \sim \tilde{p}_t \in \arg\max_q \mathbb{E}_{x\sim q}\left[\sigma_t\left(\phi_s^t(x)\right)\right] - \beta\,\mathrm{KL}\left(q \,\|\, p_1^{\theta_t}\right) \tag{12}$$

Here, the first term favors informative queries about the verifier v, and thus about the valid design space Ω^*. The KL term regularizes search toward the current generative prior, biasing exploration toward likely valid regions. The parameter β controls the exploration–prior trade-off: as $\beta \to \infty$, Eq. (12) recovers standard sampling, $x_{t+1} \sim p_1^{\theta_t}$; as $\beta \to 0$, it becomes pure uncertainty maximization, which may target remote regions where the model prior might no longer provide a useful validity bias.

To instantiate Eq. (12), we model verifier uncertainty directly in $\mathcal{Z}_s$. Let $\phi_s^t : \mathcal{X} \to \mathcal{Z}_s$ denote the representation extracted from the velocity network at noising level $s \in (0,1)$. We then view the verifier labels as noisy observations $y_t = \tilde{v}(x_t)$ of an unknown validity function over the representation space $\mathcal{Z}_s$. To represent the uncertainty about this function, we use a linear kernel over the learned representation space: $k_{\phi_s^t}(x, x') := \langle \phi_s^t(x), \phi_s^t(x') \rangle$. Then, for a set of queried designs $x_1, \ldots, x_t$, let $X_t := (x_1, \ldots, x_t)$, let $K_{t,\phi_{s,t}} \in \mathbb{R}^{t\times t}$ be the kernel matrix with entries $(K_t)_{ij} = k_{\phi_s^t}(x_i, x_j)$, and write $k_{\phi_s^t}(x, X_t) := \left(k_{\phi_s^t}(x, x_1), \ldots, k_{\phi_s^t}(x, x_t)\right)$. We then express our uncertainty about the unknown verifier via the following closed-form expression:

$$\sigma_t^2(x) = k_{\phi_s^t}(x,x) - k_{\phi_s^t}(x, X_t)\left(K_{t,\phi_s^t} + \lambda I\right)^{-1} k_{\phi_s^t}(X_t, x) \tag{13}$$

$$\sigma_t(x) := \sqrt{\sigma_t^2(x)}.$$

This corresponds to the posterior variance in Bayesian linear regression under Gaussian observations.

Algorithm 1 Active Flow Expansion (ACTFLOW)

Require: Flow model μ_{θ_0}, black-box verifier $\tilde{v}$, representation time-step $s \in (0,1)$, iterations T

$\mathcal{D}_0 \leftarrow \emptyset$
for $t = 0, 1, \ldots, T-1$ **do**
  Update surrogate uncertainty σ_t from $\mathcal{D}_t$
  Self-generate x_{t+1} according to:
  $$p_t \in \arg\max_q \mathbb{E}_{x\sim q}\left[\sigma_t\left(\phi_s^t(x)\right)\right] - \beta\,\mathrm{KL}\left(q \,\|\, p_1^{\theta_t}\right)$$
  Query verifier: $y_{t+1} \leftarrow \tilde{v}(x_{t+1})$
  $\mathcal{D}_{t+1} \leftarrow \mathcal{D}_t \cup \{(x_{t+1}, y_{t+1})\}$
  $\theta_{t+1} \leftarrow$ UPDATEFLOW$(\theta_t, \mathcal{D}_{t+1})$
end for
return μ_{θ_T}

Model update. The uncertainty estimator is fit on the pairs in $\mathcal{D}_t$. The flow model is then updated by replay-based continued pre-training. Let $\mathcal{D}_t^+ = \{x : (x,1) \in \mathcal{D}_t\}$ and $\mathcal{D}_t^- = \{x : (x,0) \in \mathcal{D}_t\}$ denote accepted and rejected samples, and let $U_t^\pm \subseteq \mathcal{D}_t^\pm$ be minibatches and $\widehat{L}_t^\pm(\theta)$ their respective standard flow-matching loss terms. We use the signed update $g_t = \nabla \widehat{L}_t^+(\theta_t) - \alpha_t \nabla \widehat{L}_t^-(\theta_t)$, where rejected samples are seen as an unlearning signal (Alberti et al., 2025); see Apx. C.1. In practice, our results often hold with $\alpha_t = 0$, i.e., standard flow matching (Lipman et al., 2024) on verifier-accepted samples.

Across iterations, ACTFLOW reallocates model mass from dominant pre-trained modes toward newly discovered valid regions. It remains to show that this process truly expands the model's generable set, rather than only redistributing density within it. The next section answers this question affirmatively by establishing statistical guarantees for out-of-distribution generative modeling via ACTFLOW.

4. Statistical Learning Guarantees for Out-of-Distribution Flow Modeling

We now theoretically analyze Active Flow Expansion (ACTFLOW), thereby providing a first guarantee for out-of-distribution generative modeling in terms of generable set expansion rather than distribution matching of p_{data}. Formal theorem and detailed derivations are reported in Apx. E.

As a first step, since ACTFLOW effectively performs active exploration in the learned representation $\mathcal{Z}_s$, we introduce the following notion of *generable representation set* $\Omega^{\tau}_{\pi,\phi}$ for a fixed map $\phi := \phi_s$.

Definition 2 (Generable Representation Set)**.** *Let $\phi : \mathcal{X} \to \mathcal{Z}$ be fixed, and let π be a diffusion model inducing design-space density p_1^{π} on $\mathcal{X}$. Let $p_1^{\pi,\phi}$ denote the induced density on $\mathcal{Z}$ obtained by pushing forward p_1^{π} through ϕ. For $\tau > 0$, we define its τ-level* generable representation set *as*

$$\Omega^{\tau}_{\pi,\phi} := \{z \in \mathcal{Z} \mid p_1^{\pi,\phi}(z) \geq \tau\}. \tag{14}$$

Formal feedback model: logistic bandit feedback We now formalize the verifier as a probabilistic model over $Z = \phi(X)$. At each round t, ACTFLOW generates a sample $z_t := \phi(x_t)$ and observes a binary label $y_t \in \{0, 1\}$, generated according to a logistic model:

$$\Pr[y_t = 1 \mid z_t] = s(g(z_t)) \tag{15}$$

where $g : Z \to \mathbb{R}$ is an unknown latent *validity score*, and $s : \mathbb{R} \to [0, 1]$ is the sigmoid function. We assume there exists a threshold $h \in [0, 1]$, such that we can formally define the valid design space as:

$$\Omega_\star := \{z \in Z : s(g(z)) \geq h\}. \tag{16}$$

For the sake of analysis, we assume g lies in an RKHS $\mathcal{H}_k$ with $\|g\|_k \leq B$, $\int_{\mathcal{Z}} \exp(g(z)) \leq \bar{Z}$, and is L_g-Lipschitz in a metric d over $\mathcal{Z}$. This model allows us to construct *anytime confidence sequences* for logistic prediction (Pásztor et al., 2024). Next, we model ACTFLOW's uncertainty-tilted sampling procedure.

Uncertainty-tilted sampling oracle modeling We model the inference-time uncertainty sampling in Eq. (12) as approximately maximizing verifier uncertainty within the current generable set Ω^{τ}_t for some $\tau > 0$. Concretely, at round t it returns x_t such that $z_t = \phi(x_t) \in \Omega^{\tau}_t$ satisfies:

$$\sigma_t(z_t) \;\geq\; \frac{1}{\alpha} \max_{z \in \Omega^{\tau}_t} \sigma_t(z), \qquad \alpha \geq 1. \tag{17}$$

We depart from standard bandit analyses, which assume global search over $\Omega^\star$ by instead restricting the (generative) sampler to approximate uncertainty maximization over the generable set. We suppose that the τ-level generable set of the pre-trained model contains a non-empty set S_0 of valid designs. Further regions can be reached only after intermediate expansion steps. We formalize a local expansion step via the following *one-step reachability operator over the learned representation*:

$$R_\epsilon(S) := \{z \in \mathcal{Z} : \exists z' \in S : s(g(z')) - L_s L_g d(z, z') - \epsilon \geq h\}$$

Here, $R_\epsilon(S)$ contains the representations whose validity is certifiable from S up to accuracy ϵ. Let $R^H_\epsilon(S_0)$ denote its H-fold recursive application starting from S_0. This enables viewing of ACTFLOW as executing a local-to-global expansion process toward the reachable valid set $R^H_\epsilon(S_0)$.

We can now present the main theorem, presented formally in Theorem E.5. It allows to state set-theoretic guarantees for out-of-distribution generative modeling.

(Informal) Assumptions:

- **(Well-specified verifier).** Verifier labels follow a calibrated logistic model in the learned representation space, $\mathbb{P}(y = 1 \mid z) = s(g(z))$, where the latent validity score g has bounded RKHS norm and is Lipschitz continuous in the learned representation.
- **(Approximate and local uncertainty sampling).** At each round, the sampler approximately maximizes verifier uncertainty over the current generable set Ω^{τ}_t, as by Eq. (17).
- **(EBM generative update).** We employ an energy-based model (EBM) abstraction, detailed in Assump. E.1, to abstract flow model endpoint density updates as learning an implicit energy functional over the flow-learned representation Z. This captures that flow models preserve high density on regions that our verifier regards as valid, and maintain low density on points far from this set.

Theorem 4.1 ((Informal) Generable representation set covers reachable set)**.** *Fix $\epsilon > 0$ and an integer $H \geq 1$. Let τ and T^* satisfy*

$$\tau \leq \frac{h}{(1-h)\bar{Z}} \quad \text{and} \quad T^\star \;\gtrsim\; \left(\frac{\alpha \gamma^{\Omega_\star}_{HT^*}}{\epsilon}\right)^2, \tag{18}$$

with up to problem-dependent constants in $\gtrsim$. Define the maximum information gain from t samples by

$$\gamma^A_t \triangleq \max_{z_1,\ldots,z_t \in A} \frac{1}{2} \log\det\Big(I_t + (\lambda\kappa)^{-1} K_t\Big)$$

By running ACTFLOW*, it holds with probability at least $1 - \delta$ that after $T^\star$ verified samples,*

$$R^H_\epsilon(S_0) \subseteq \Omega^{\tau}_{T^\star}. \tag{19}$$

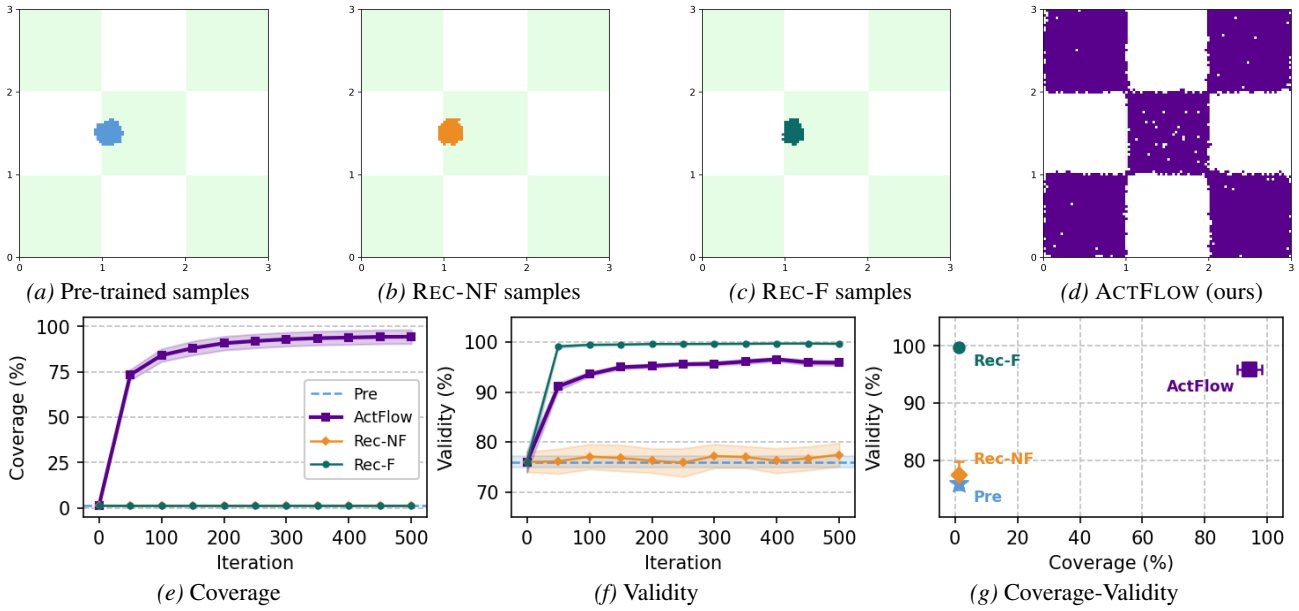


Figure 2. (2a) The valid design space (green) and the estimated generable set ($\tau = 0.01$) of a pre-trained model (blue). (2a-2d) REC-NF and REC-F fail to expand the pre-trained model's generable set, while ACTFLOW is able to discover and expand over the entire valid design space. (2e-2g) ACTFLOW increases both model coverage (1.16% to 94.27%) and validity (76% to 95.9%), REC-NF fails at increasing both, while REC-F increases validity while even decreasing coverage (1.16% to 1.1%).

Theorem 4.1 is stated in the learned representation space Z. Under a map ϕ with measure-preserving regularity conditions stated in Assumption E.4, the same reachability guarantee transfers to the design-space generable set. For $\tau_X > 0$, define the design-space generable set at round t by

$$\Omega_t^{X,\tau_X} := \{x \in \mathcal{X} : \; p_1^{\pi_t}(x) \geq \tau_X\}. \tag{20}$$

We also define the induced one-step reachability operator on $\mathcal{X}$ by

$$R_\epsilon^{X,\phi}(A) := \Big\{x \in \mathcal{X} : \exists x' \in A \text{ s.t.} \tag{21}$$
$$s\big(g(\phi(x'))\big) - L_s L_g \, d\big(\phi(x), \phi(x')\big) - \epsilon \geq h\Big\},$$

for any $A \subseteq \mathcal{X}$, and let $\big(R_\epsilon^{X,\phi}\big)^H(A)$ denote its H-fold iterate.

Corollary 4.2 (Design-space coverage of the induced reachable valid set)**.** *Assume the conditions of Theorem 4.1 and Assumption E.4, with* $j_{\min} := \inf_{x\in\mathcal{X}} |\det J_\phi(x)| > 0$. *Let* $S_0^X := \phi^{-1}(S_0)$ *and* $\tau_X := j_{\min}\tau$. *Then, after the same number* $T^\star$ *of verified samples as in Theorem 4.1, with probability at least* $1-\delta$,

$$\big(R_\epsilon^{X,\phi}\big)^H(S_0^X) \subseteq \Omega_{T^\star}^{X,\tau_X}. \tag{22}$$

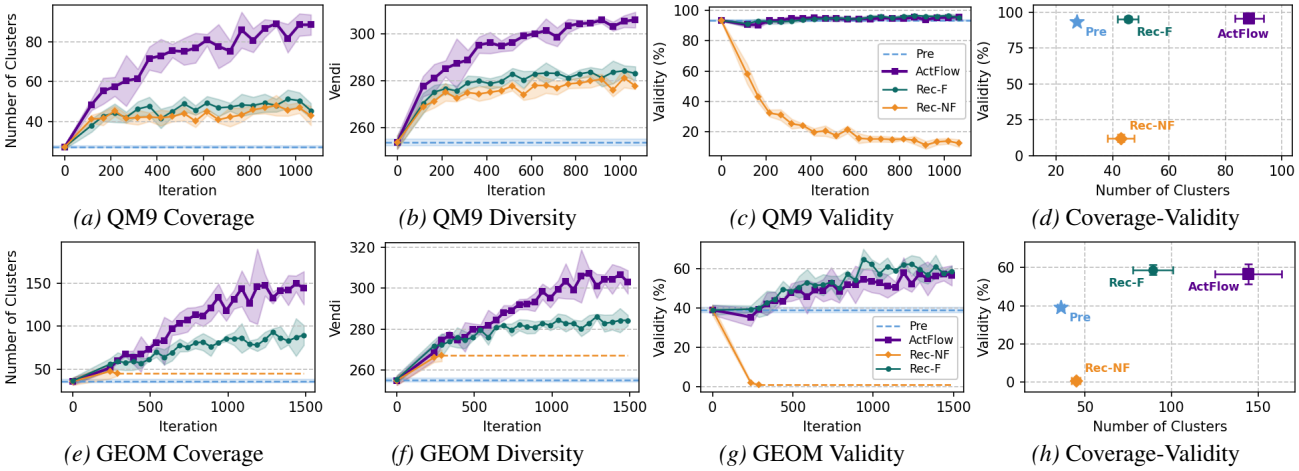
From reachable-set to full coverage. The guarantee is reachability-based: if $\Omega^\star$ contains components that are disconnected from S_0^X in the learned geometry, local expansion provides no guarantee of covering those components. Conversely, if every $x \in \Omega^\star$ can be connected to S_0^X by at most H local valid expansions in representation space, then Corollary E.9 yields full valid-space coverage,

$$\Omega^\star \subseteq \Omega_{T^\star}^{X,\tau_X}.$$

Theorem 4.1 shows that ACTFLOW expands the model's τ-generable set for any sufficiently small τ, effectively describing finite-sample coverage. Corollary E.8 controls model validity, i.e., how tight the expanded generable set is w.r.t. $\Omega^\star$ – the opposite set-inequality for *generable set expansion*.

5. Experimental Evaluation of ActFlow on Molecules, Peptides, and Proteins

We evaluate ACTFLOW for expanding the valid generable set of pre-trained generative models, and compare it against widely adopted self-generation baselines: continued pre-training on unfiltered model samples (REC-NF) (Shumailov et al., 2024; Alemohammad et al., 2023) and on verifier-filtered samples (REC-F) (Dong et al., 2023; Gulcehre et al., 2023). We propose two types of experiments: (i) an illustrative visually interpretable setting, and (ii) high-dimensional biochemical design tasks over molecules, therapeutic peptides, and protein sequences. Since standard generative modeling evaluation metrics are misaligned with OOD generative modeling, we employ the OOD modeling criteria presented in Apx. G.1, namely *coverage* via number of valid clusters, *diversity* via Vendi (Friedman & Dieng, 2022), and overall model validity. Additional experimental



(a) QM9 Coverage *(b)* QM9 Diversity *(c)* QM9 Validity *(d)* Coverage-Validity

(e) GEOM Coverage *(f)* GEOM Diversity *(g)* GEOM Validity *(h)* Coverage-Validity

Figure 3. (3a-3d) Molecular design on QM9 (Ramakrishnan et al., 2014). ACTFLOW expands valid coverage substantially more than REC-NF and REC-F, reaching 88.40 valid clusters versus 42.80 and 45.40 respectively (Fig. 3a). It also achieves the highest diversity (3b) while preserving high validity (3d). (3e-3h) Design of drug-like molecules on GEOM-Drugs (Axelrod & Gomez-Bombarelli, 2022). ACTFLOW increases the pre-trained model coverage, diversity, and validity, vastly outperforming both REC-F and REC-NF baselines.

details are provided in Apx. G.6.

Illustrative Visually Interpretable Setting. We first evaluate ACTFLOW in a two-dimensional illustrative setting with valid design space shown in green in Fig. 2a–2d, and visualize the models' generable sets via discretization. All methods run for $T = 500$ iterations and generate $B = 64$ samples per iteration. ACTFLOW uses $\alpha_t = 0.005$, $\beta = 1/13$, and flow representation at timestep $s = 0.9$. Fig. 2a–2d show that ACTFLOW (violet) expands the pre-trained generable set ($\tau = 0.01$, Fig. 2a) to near-optimally cover the valid design space (Fig. 2c), whereas both baselines fail to expand it. Fig. 2e–2g show that ACTFLOW increases both coverage, from 1.16% to 94.27%, and validity, from 76.00% to 95.89%. By contrast, REC-NF leaves coverage unchanged and barely improves validity, while REC-F attains higher validity but decreases coverage. Thus, even in this simple two-dimensional setting, ACTFLOW is the only method that achieves OOD generable-set expansion, substantially improving both coverage and validity, expanding through sparse directions (corners) to new valid regions.

Molecular design on QM9. We evaluate ACTFLOW on FlowMol Gaussian (Dunn & Koes, 2024), pre-trained on QM9 (Ramakrishnan et al., 2014). All methods run for 1000 iterations after 66 initial iterations without fine-tuning. ACTFLOW uses the flow representation at timestep $s = 0.9$ and $\beta = 1/10$. Fig. 3a–3d show that, relative to the pre-trained model, ACTFLOW substantially expands valid molecular coverage, reaching 88.40 valid clusters, compared to 45.40 for REC-F and 42.80 for REC-NF. ACTFLOW also achieves the highest diversity, with Vendi 306.08, while preserving high validity (95.90%). In contrast, REC-F preserves validity (95.24%) but expands significantly less, whereas REC-NF severely degrades validity (12.26%). Thus, on QM9, ACTFLOW significantly outperforms both recursive sampling baselines, achieving a significantly stronger combination of valid coverage, diversity, and validity.

Molecular design on GEOM-Drugs. We further evaluate ACTFLOW on FlowMol Gaussian (Dunn & Koes, 2024), pre-trained on GEOM-Drugs (Axelrod & Gomez-Bombarelli, 2022), a substantially larger, and more chemically relevant dataset of drug-like molecules. All methods use 2000 fine-tuning steps after a warm-up period, where 4096 samples are acquired. ACTFLOW uses flow representation timestep $s = 0.8$, $\beta = 1/7$, $\alpha_t = 0$. Fig. 3e–3h show that ACTFLOW expands valid molecular coverage substantially beyond the baselines, reaching 144.3 valid clusters, compared to 89.33 for REC-F and 44.67 for REC-NF. ACTFLOW also achieves the highest diversity, with Vendi 303.1, while maintaining comparable validity to REC-F (56.6% versus 58.83%). In contrast, REC-F expands substantially less, whereas REC-NF collapses in validity, as reported in Fig. 3h. Thus, on GEOM-Drugs, ACTFLOW provides a stronger expansion profile than both baselines, ultimately increasing the pre-trained model coverage by 144.3% and its validity by 56.6%.

Therapeutic Peptide Design We assess ACTFLOW on biological sequence generation via discrete diffusion models, as detailed in Apx. F.3. We consider the task of therapeutic peptide design (Wang et al., 2022), and employ the pre-trained SMILES discrete diffusion model from PepTune (Tang et al., 2025). As shown in Table 2 and Fig 4, ACTFLOW significantly increases the number of clusters (i.e., coverage) from 133.3 to 197, while further increasing validity, determined via the SMILES2PEPTIDE verifier (Tang et al., 2025), from 31.9% to 52.0%. Similarly,

ACTFLOW achieves the highest diversity of 7.45 Vendi score over PeptideCLM embeddings (Feller & Wilke, 2025). On the contrary, REC-F decreases the pre-trained model coverage to 45 clusters, while increasing validity, and REC-NF collapses entirely to 0.0% validity, as a single misplaced token can result in an invalid peptide – thus inducing no valid clusters. These results show that ACTFLOW's strong empirical performance extends beyond continuous flows to discrete diffusion models (see Apx. F.3), significantly increasing jointly model coverage, diversity, and validity, and outperforming baselines.

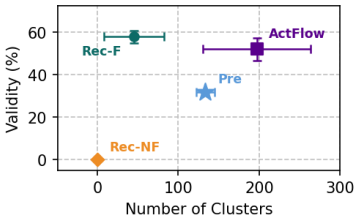


Figure 4. Peptides coverage-validity.

Protein Sequence Design. We finally evaluate ACTFLOW on protein sequence design, via a continuous ESM diffusion model from SGPO (Yang et al., 2025), pre-trained on the CreiLOV fluorescence dataset (Chen et al., 2023). We use 512 iterations with 1000 fine-tuning steps each. ACTFLOW employs flow representation timestep at $t = 0.8$. ACTFLOW substantially expands valid coverage, computed over token-level ESM embeddings, increasing the number of clusters from 66.50 to 102.75, compared to 63.75 for REC-NF and 49.50 for REC-F, which lead to significant decrease in both coverage and validity metrics, see Table 2 and Fig. 5. ACTFLOW also achieves the highest diversity, with Vendi 42.14 vs 12.87 for the pre-trained model and 11.85/11.67 for the baselines, while improving validity from 70.81% to 83.74%. In contrast, REC-F reaches higher validity (88.12%) but collapses in coverage and diversity; REC-NF reduces coverage and diversity relative to the pre-trained model. ACTFLOW achieves the strongest coverage–diversity–validity profile (see Fig. 5), significantly expanding the pre-trained protein sequence model to new valid regions of the design space.

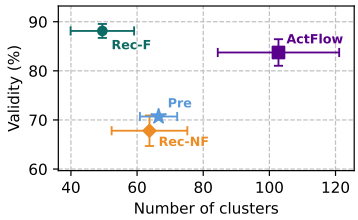


Figure 5. Proteins coverage-validity.

6. Related Work

Diffusion and flow model reward adaptation for generative optimization. Several works adapt pre-trained diffusion and flow models for reward maximization, either via fine-tuning (e.g., Uehara et al., 2024; De Santi et al., 2025b; Liu et al., 2025) or inference-time sampling (e.g., Uehara et al., 2025; Perez Jensen et al., 2026; Uehara et al., 2025). Without injecting further validity information, these methods remain constrained by the pre-trained model coverage, leading to *over-optimization*, i.e., invalid samples, when steering the model excessively beyond the training distribution (Gutjahr et al., 2025). In this work, we formalize this limitation through the notion of *generable set*, and introduce ACTFLOW, a task-agnostic continued pre-training method that expands coverage to new valid regions of the design space.

Synthetic data generation for diffusion model self-adaptation Generative priors can provide useful synthetic data for downstream learning (e.g., Azizi et al., 2023; Nguyen et al., 2025), but recursively training on generated data can induce model collapse (Shumailov et al., 2024; Alemohammad et al., 2023). For generative model adaptation, prior work uses synthetic data for negative guidance (Alemohammad et al., 2024) or conservative verifier-free self-play (Yuan et al., 2024), with the primary goal of improving in-distribution sample quality (e.g., Alemohammad et al., 2024). In contrast, we study flow adaptation for out-of-distribution modeling: rather than further refining high-density modes, the goal is to reallocate mass toward newly verified valid regions beyond the training distribution. To our knowledge, ACTFLOW is the first theory-backed synthetic continued-pretraining method for out-of-distribution generative modeling.

Diffusion and flow based design space exploration Recent works introduce scalable methods for diffusion- and flow-based design space exploration via entropy maximization (e.g., De Santi et al., 2025a;b) or approximations (e.g., Celik et al., 2025). A more recent line assumes access to a known differentiable verifier and uses it to rebalance model density toward valid regions via verifier-constrained entropy maximization (De Santi et al., 2026). In contrast, we assume access only to black-box, unknown verifiers, covering highly-relevant settings in science, where feedback is experimental and/or non-differentiable. Moreover, ACTFLOW expands valid coverage through directed synthetic data generation rather than verifier-constrained optimization.

Safe (active) exploration theory Safe exploration studies learning under unknown safety constraints. Early work uses Gaussian-process confidence bounds to restrict exploration to points certified above a safety threshold (Sui et al., 2015; 2018). Subsequent safe-RL methods extend this principle to MDPs and control (e.g., Turchetta et al., 2016; Berkenkamp et al., 2017; Koller et al., 2018). We bridge *safety* in safe exploration with *validity* in generative modeling: expanding a generable set amounts to discovering new valid (i.e., safe) regions. In scientific discovery, invalid samples lead to rejected verifier queries, not unsafe actions in safety-critical systems; thus, ACTFLOW does not perform safe-set constrained planning. Rather, we employ this viewpoint to provide a first-of-its-kind reachability-based theory of out-of-distribution flow modeling.

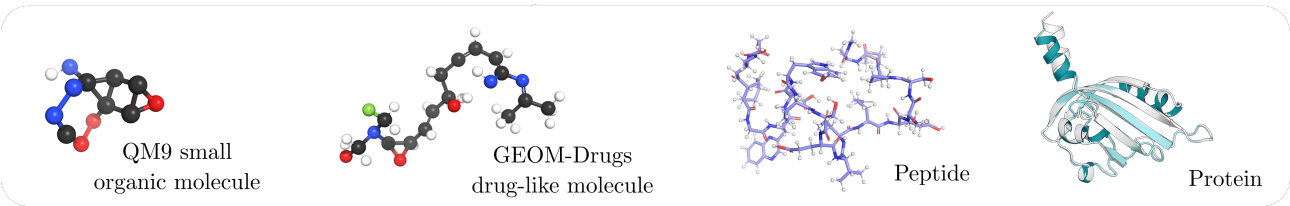


Figure 6. 3D structures directly generated or string-converted by models expanded via ACTFLOW.

	Therapeutic Peptide Design				Protein Sequence Design			
Method	**Coverage** ↑	**Diversity** ↑	**FID**	**Validity** (%) ↑	**Coverage** ↑	**Diversity** ↑	**FID**	**Validity** (%) ↑
Pre	133.30 ± 11.1	5.70 ± 0.11	0.00 ± 0.00	31.92 ± 0.77	66.50 ± 5.63	12.87 ± 0.63	0.05 ± 0.01	70.81 ± 1.12
REC-NF	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	63.75 ± 11.45	11.85 ± 0.34	0.27 ± 0.06	67.81 ± 3.11
REC-F	45.00 ± 37.22	6.03 ± 1.19	4.99 ± 1.53	57.84 ± 2.97	49.50 ± 9.60	11.67 ± 0.40	0.25 ± 0.04	88.12 ± 1.42
ACTFLOW	197.00 ± 66.25	7.45 ± 6.17	4.49 ± 6.44	52.02 ± 5.34	102.75 ± 18.36	42.14 ± 10.85	5.45 ± 3.34	83.74 ± 2.70

Table 2. ACTFLOW significantly expands valid model coverage (i.e., number of valid clusters) and diversity (i.e., Vendi (Friedman & Dieng, 2022)), while also increasing validity across therapeutic peptide and protein sequence design tasks. We report in red models where coverage or diversity decreased from initial model. These results show that ACTFLOW vastly outperforms widely adopted synthetic pre-training baselines.

7. Conclusion

We depart from data distribution matching and formulate *out-of-distribution flow modeling* via *generable set expansion*, i.e., expanding the valid region of design space a model samples with non-negligible probability. We introduced ACTFLOW, a continued-pretraining scheme that uses verifier feedback on self-generated data to actively expand in the learned flow representation – leading to first-of-their-kind reachability-based guarantees for out-of-distribution flow modeling. Across small organic molecules, drug-like molecules, therapeutic peptides, and protein sequences, ACTFLOW consistently improves coverage, diversity, and validity over widely adopted recursive self-generation baselines. As for limitations, while our framework allows for task-agnostic expansions toward valid, previously inaccessible regions where new-to-nature discoveries may reside – future work will need to assess whether this form of exploration yields concrete gains (i.e., discoveries) in specific real-world applications.

Impact Statement

This paper presents work whose goal is to advance the field of Generative Modeling. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

Acknowledgments

This publication was made possible by the ETH AI Center doctoral fellowship to Riccardo De Santi. The project has received funding from the Swiss National Science Foundation under NCCR Catalysis (grant number 180544 and 225147) and NCCR Automation (grant agreement 51NF40 180545), National Centres of Competence in Research funded by the Swiss National Science Foundation. This work was supported by an ETH Zurich Research Grant. This research was also supported by a grant from the High-throughput Institute for Discovery (HIT-ID) at the University of Pennsylvania to the laboratory of Pranam Chatterjee.

References

Alberti, S., Hasanaliyev, K., Shah, M., and Ermon, S. Data unlearning in diffusion models. *arXiv preprint arXiv:2503.01034*, 2025.

Alemohammad, S., Casco-Rodriguez, J., Luzi, L., Humayun, A. I., Babaei, H., LeJeune, D., Siahkoohi, A., and Baraniuk, R. Self-consuming generative models go mad. In *The Twelfth International Conference on Learning Representations*, 2023.

Alemohammad, S., Humayun, A. I., Agarwal, S., Collomosse, J., and Baraniuk, R. Self-improving diffusion models with synthetic data. *arXiv preprint arXiv:2408.16333*, 2024.

Axelrod, S. and Gomez-Bombarelli, R. Geom, energy-annotated molecular conformations for property prediction and molecular generation. *Scientific Data*, 9(1):185, 2022.

Azizi, S., Kornblith, S., Saharia, C., Norouzi, M., and Fleet, D. J. Synthetic data from diffusion models improves

imagenet classification. *arXiv preprint arXiv:2304.08466*, 2023.

Berkenkamp, F., Turchetta, M., Schoellig, A., and Krause, A. Safe model-based reinforcement learning with stability guarantees. *Advances in neural information processing systems*, 30, 2017.

Boucheron, S., Lugosi, G., and Massart, P. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

Celik, O., Li, Z., Blessing, D., Li, G., Palenicek, D., Peters, J., Chalvatzaki, G., and Neumann, G. Dime: Diffusion-based maximum entropy reinforcement learning. *arXiv preprint arXiv:2502.02316*, 2025.

Chen, Y., Hu, R., Li, K., Zhang, Y., Fu, L., Zhang, J., and Si, T. Deep mutational scanning of an oxygen-independent fluorescent protein creilov for comprehensive profiling of mutational and epistatic effects. *ACS Synthetic Biology*, 12(5):1461–1473, 2023.

Chi, C., Xu, Z., Feng, S., Cousineau, E., Du, Y., Burchfiel, B., Tedrake, R., and Song, S. Diffusion policy: Visuomotor policy learning via action diffusion. *The International Journal of Robotics Research*, 44(10-11): 1684–1704, 2025.

Corso, G., Stärk, H., Jing, B., Barzilay, R., and Jaakkola, T. Diffdock: Diffusion steps, twists, and turns for molecular docking. *arXiv preprint arXiv:2210.01776*, 2022.

De Santi, R., Vlastelica, M., Hsieh, Y.-P., Shen, Z., He, N., and Krause, A. Provable maximum entropy manifold exploration via diffusion models. In *International Conference on Machine Learning*, 2025a.

De Santi, R., Vlastelica, M., Hsieh, Y.-P., Shen, Z., He, N., and Krause, A. Flow density control: Generative optimization beyond entropy-regularized fine-tuning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025b.

De Santi, R., Protopapas, K., Hsieh, Y.-P., and Krause, A. Verifier-constrained flow expansion for discovery beyond the data. In *International Conference on Learning Representations (ICLR)*, April 2026.

Dong, H., Xiong, W., Goyal, D., Zhang, Y., Chow, W., Pan, R., Diao, S., Zhang, J., Shum, K., and Zhang, T. Raft: Reward ranked finetuning for generative foundation model alignment. *arXiv preprint arXiv:2304.06767*, 2023.

Dunn, I. and Koes, D. R. Mixed continuous and categorical flow matching for 3d de novo molecule generation. *ArXiv*, pp. arXiv–2404, 2024.

Feller, A. L. and Wilke, C. O. Peptide-aware chemical language model successfully predicts membrane diffusion of cyclic peptides. *Journal of chemical information and modeling*, 65(2):571–579, 2025.

Friedman, D. and Dieng, A. B. The vendi score: A diversity evaluation metric for machine learning. *arXiv preprint arXiv:2210.02410*, 2022.

Gulcehre, C., Paine, T. L., Srinivasan, S., Konyushkova, K., Weerts, L., Sharma, A., Siddhant, A., Ahern, A., Wang, M., Gu, C., et al. Reinforced self-training (rest) for language modeling. *arXiv preprint arXiv:2308.08998*, 2023.

Guo, J. and Schwaller, P. It takes two to tango: Directly optimizing for constrained synthesizability in generative molecular design. *arXiv preprint arXiv:2410.11527*, 2024.

Gutjahr, S., De Santi, R., Schaufelberger, L., Jorner, K., and Krause, A. Constrained molecular generation via sequential flow model fine-tuning. In *ICML 2025 Generative AI and Biology (GenBio) Workshop*, 2025.

Halgren, T. A. Merck molecular force field. i. basis, form, scope, parameterization, and performance of mmff94. *Journal of computational chemistry*, 17(5-6):490–519, 1996.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

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

Hoogeboom, E., Satorras, V. G., Vignac, C., and Welling, M. Equivariant diffusion for molecule generation in 3d. In *International conference on machine learning*, pp. 8867–8887. PMLR, 2022.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. Highly accurate protein structure prediction with alphafold. *nature*, 596(7873):583–589, 2021.

Koller, T., Berkenkamp, F., Turchetta, M., and Krause, A. Learning-based model predictive control for safe exploration. In *2018 IEEE conference on decision and control (CDC)*, pp. 6059–6066. IEEE, 2018.

Laurent, B. and Massart, P. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302–1338, 2000.

Li, X. and Fourches, D. Smiles pair encoding: a data-driven substructure tokenization algorithm for deep learning. *Journal of chemical information and modeling*, 61(4): 1560–1569, 2021.

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., dos Santos Costa, A., Fazel-Zarandi, M., Sercu, T., Candido, S., et al. Language models of protein sequences at the scale of evolution enable accurate structure prediction. *bioRxiv*, 2022.

Lipman, Y., Chen, R. T., Ben-Hamu, H., Nickel, M., and Le, M. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Lipman, Y., Havasi, M., Holderrieth, P., Shaul, N., Le, M., Karrer, B., Chen, R. T., Lopez-Paz, D., Ben-Hamu, H., and Gat, I. Flow matching guide and code. *arXiv preprint arXiv:2412.06264*, 2024.

Liu, J., Liu, G., Liang, J., Li, Y., Liu, J., Wang, X., Wan, P., Zhang, D., and Ouyang, W. Flow-grpo: Training flow matching models via online rl. *arXiv preprint arXiv:2505.05470*, 2025.

Lou, A., Meng, C., and Ermon, S. Discrete diffusion modeling by estimating the ratios of the data distribution. *arXiv preprint arXiv:2310.16834*, 2023.

Nguyen, D., Li, J., Zheng, J., and Mirzasoleiman, B. Do we need all the synthetic data? targeted image augmentation via diffusion models. In *The Fourteenth International Conference on Learning Representations*, 2025.

Ou, J., Nie, S., Xue, K., Zhu, F., Sun, J., Li, Z., and Li, C. Your absorbing discrete diffusion secretly models the conditional distributions of clean data. *arXiv preprint arXiv:2406.03736*, 2024.

Pásztor, B., Kassraie, P., and Krause, A. Bandits with preference feedback: A stackelberg game perspective. *Advances in Neural Information Processing Systems*, 37: 11997–12034, 2024.

Pearce, T., Rashid, T., Kanervisto, A., Bignell, D., Sun, M., Georgescu, R., Macua, S. V., Tan, S. Z., Momennejad, I., Hofmann, K., et al. Imitating human behaviour with diffusion models. *arXiv preprint arXiv:2301.10677*, 2023.

Perez Jensen, C., Schaufelberger, L., Santi, R. D., Jorner, K., and Krause, A. Value matching: Scalable and gradient-free reward-guided flow adaptation. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=7iXt44Actj.

Ramakrishnan, R., Dral, P. O., Rupp, M., and Von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific data*, 1(1):1–7, 2014.

Rector-Brooks, J., Lambert, T., Skreta, M., Roth, D., Long, Y., Li, Z.-Q., Zhang, X., Cretu, M., Li, F.-Z., Ganapathy, T., et al. General multimodal protein design enables dna-encoding of chemistry. *arXiv preprint arXiv:2604.05181*, 2026.

Sahoo, S., Arriola, M., Schiff, Y., Gokaslan, A., Marroquin, E., Chiu, J., Rush, A., and Kuleshov, V. Simple and effective masked diffusion language models. *Advances in Neural Information Processing Systems*, 37:130136–130184, 2024.

Schölkopf, B., Herbrich, R., and Smola, A. J. A generalized representer theorem. In *International conference on computational learning theory*, pp. 416–426. Springer, 2001.

Shi, J., Han, K., Wang, Z., Doucet, A., and Titsias, M. Simplified and generalized masked diffusion for discrete data. *Advances in neural information processing systems*, 37:103131–103167, 2024.

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., and Gal, Y. Ai models collapse when trained on recursively generated data. *Nature*, 631(8022):755–759, 2024.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., and Ganguli, S. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pp. 2256–2265. PMLR, 2015.

Song, Y. and Ermon, S. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.

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

Sui, Y., Gotovos, A., Burdick, J., and Krause, A. Safe exploration for optimization with gaussian processes. In *International conference on machine learning*, pp. 997–1005. PMLR, 2015.

Sui, Y., Zhuang, V., Burdick, J., and Yue, Y. Stagewise safe bayesian optimization with gaussian processes. In *International conference on machine learning*, pp. 4781–4789. PMLR, 2018.

Tang, S., Zhang, Y., and Chatterjee, P. Peptune: De novo generation of therapeutic peptides with multi-objective-guided discrete diffusion. *42nd International Conference on Machine Learning*, 2025.

Turchetta, M., Berkenkamp, F., and Krause, A. Safe exploration in finite markov decision processes with gaussian processes. *Advances in neural information processing systems*, 29, 2016.

Uehara, M., Zhao, Y., Biancalani, T., and Levine, S. Understanding reinforcement learning-based fine-tuning of diffusion models: A tutorial and review, 2024. URL https://arxiv.org/abs/2407.13734.

Uehara, M., Zhao, Y., Wang, C., Li, X., Regev, A., Levine, S., and Biancalani, T. Inference-time alignment in diffusion models with reward-guided generation: Tutorial and review. *arXiv preprint arXiv:2501.09685*, 2025.

Wang, L., Wang, N., Zhang, W., Cheng, X., Yan, Z., Shao, G., Wang, X., Wang, R., and Fu, C. Therapeutic peptides: current applications and future directions. *Signal transduction and targeted therapy*, 7(1):48, 2022.

Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., Eisenach, H. E., Ahern, W., Borst, A. J., Ragotte, R. J., Milles, L. F., et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976): 1089–1100, 2023.

Weininger, D. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.*, 28(1):31–36, 1988.

Yang, J., Chu, W., Khalil, D., Astudillo, R., Wittmann, B. J., Arnold, F. H., and Yue, Y. Steering generative models with experimental data for protein fitness optimization, 2025. URL https://arxiv.org/abs/2505.15093.

Yuan, H., Chen, Z., Ji, K., and Gu, Q. Self-play fine-tuning of diffusion models for text-to-image generation. *Advances in Neural Information Processing Systems*, 37: 73366–73398, 2024.

Zheng, K., Chen, Y., Mao, H., Liu, M.-Y., Zhu, J., and Zhang, Q. Masked diffusion models are secretly time-agnostic masked models and exploit inaccurate categorical sampling. *arXiv preprint arXiv:2409.02908*, 2024.

A. Central Limitation of Standard Pre-training with an Imperfect Model

In the main text, Eq. (7) is stated for clarity under the idealization that the pre-trained model generates only valid designs. In practice, this need not hold: the generable set may contain invalid designs, so that

$$\Omega^{\tau}_{\theta^{\mathrm{pre}}} \cap (\mathcal{X} \setminus \Omega^{\star}) \neq \emptyset.$$

This does not change the central limitation. One simply applies the coverage statement to the *valid part* of the generable set,

$$\Omega^{\tau,\mathrm{val}}_{\theta^{\mathrm{pre}}} := \Omega^{\tau}_{\theta^{\mathrm{pre}}} \cap \Omega^{\star} \subseteq \Omega^{\star}.$$

The realistic limitation is therefore

$$\Omega^{\tau,\mathrm{val}}_{\theta^{\mathrm{pre}}} \subset \Omega^{\star}, \qquad \mathrm{Vol}\Big(\Omega^{\tau,\mathrm{val}}_{\theta^{\mathrm{pre}}}\Big) \ll \mathrm{Vol}(\Omega^{\star}),$$

i.e., the pre-trained model covers only a small fraction of the valid design space with non-negligible probability, even if it also assigns mass to invalid regions.

This is precisely the setting addressed by ActFlow. The algorithm does not require the initial generable set to be valid; it only requires verifier feedback to identify which generated samples are valid and to adapt the model toward newly verified valid regions. Thus, generable set expansion should be understood as expansion of the valid, verifier-certified portion of the model's coverage. Moreover, when the initial model has partial validity, reallocating mass toward newly verified valid regions can also increase the model's overall validity, as we observe in our experiments, where the pre-trained models are imperfect and ActFlow systematically improves not only coverage and diversity, but typically increases substantially validity as well.

Volume over the valid design space. Throughout, $\mathrm{Vol}(\cdot)$ denotes volume with respect to a domain-appropriate reference measure on the valid design space $\Omega^{\star}$. Formally, let $\nu^{\star}$ be a reference measure supported on $\Omega^{\star}$: Lebesgue measure when $\Omega^{\star}$ is full-dimensional, intrinsic Hausdorff measure when $\Omega^{\star}$ is a lower-dimensional manifold, and counting measure in discrete design spaces. For any $A \subseteq \mathcal{X}$, we write

$$\mathrm{Vol}(A) := \nu^{\star}(A \cap \Omega^{\star}).$$

Thus, the central limitation in Eq. (7) measures how much of the valid design space is covered by the model's τ-generable set, rather than ambient Lebesgue volume in $\mathcal{X}$. This avoids degeneracies in settings where valid designs lie on lower-dimensional manifolds or discrete spaces.

B. A Warm-Up Analysis of Failure Modes of Expansion via Synthetic Data

B.1. Compact Analysis

Warm Up Gaussian Setting. A widely adopted approach to synthetic data generation is recursive sampling with closed-loop verifier filtering. The purpose of this section is *not* to establish a general impossibility result for this paradigm, but to use a minimal analytical model to isolate possible failure modes relevant to generable-set expansion. To this end, we introduce a simple probabilistic notion of the *generable valid frontier*: valid samples that remain reliably generable under the current model, yet lie away from its dominant modes. Such points form a natural abstraction of expansion-enabling samples. The analysis then shows that standard recursive self-generation schemes may be unlikely to produce them even in a simple low-dimensional setting, thereby shedding light on possible failure modes of data generation schemes and motivating algorithmic desiderata for expansion. We report derivations in Apx. B.2.

A Gaussian abstraction of pre-training. We consider a design space $\mathcal{X} = \mathbb{R}^d$. Let $U \subset \mathbb{R}^d$ be a k-dimensional linear subspace, let $m := d-k$, and write $\mathbb{R}^d = U \oplus U^{\perp}$. We interpret U as a low-dimensional region well captured by pre-training, and $U^{\perp}$ as orthogonal directions along which expansion beyond the dominant pre-trained modes could occur. For $x \in \mathbb{R}^d$, write $x = x_U + x_{\perp}$ with $x_U = \Pi_U x$ and $x_{\perp} = \Pi_{U^{\perp}} x$. We model the pre-trained generator p^{θ_0} by the anisotropic Gaussian

$$X \sim p^{\theta_0} =: p_0 = \mathcal{N}\big(0,\, I_U \oplus \sigma^2 I_{U^{\perp}}\big), \qquad 0 < \sigma \ll 1.$$

Thus, the model is spread along U, while it is sharply concentrated around U in the orthogonal directions. For a density threshold $\tau > 0$, the corresponding generable set is

$$\Omega_{\theta_0}^{\tau} = \left\{x \in \mathbb{R}^d : \|x_U\|_2^2 + \sigma^{-2}\|x_\perp\|_2^2 \le r_\tau^2\right\}, \qquad r_\tau^2 := 2\log\left(\frac{(2\pi)^{-d/2}\sigma^{-m}}{\tau}\right),$$

namely an ellipsoid elongated along U and very thin along $U^\perp$.

A probabilistic notion of generable valid frontier. To reason about samples that may support expansion, we isolate a shell inside $\Omega_{\theta_0}^{\tau}$ that is close to its orthogonal boundary. Fix $R_U > 0$ and radii $0 < \rho_- < \rho_+$ such that $R_U < r_\tau$ and $\rho_+ < \sigma\sqrt{r_\tau^2 - R_U^2}$. We denote the *generable frontier* by

$$\mathcal{F}_{\text{frontier}} := \left\{x \in \mathbb{R}^d : \|x_U\|_2 \le R_U,\ \rho_- \le \|x_\perp\|_2 \le \rho_+\right\}.$$

By construction, $\mathcal{F}_{\text{frontier}} \subseteq \Omega_{\theta_0}^{\tau}$. Thus, points in $\mathcal{F}_{\text{frontier}}$ remain reliably generable under the current model, since they lie inside its generable set, while already exhibiting a substantial orthogonal deviation from the model dominant modes. Next, fix a unit vector $u_\star \in U^\perp$ and an opening angle $\phi \in (0, \pi/2)$. We define the cone of *valid directions of expansion*

$$C_\phi(u_\star) := \left\{y \in U^\perp \setminus \{0\} : \left\langle \frac{y}{\|y\|_2}, u_\star \right\rangle \ge \cos\phi\right\}.$$

We call *generable valid frontier* the set

$$\mathcal{V}_{\text{frontier}} := \left\{x \in \mathcal{F}_{\text{frontier}} : x_\perp \in C_\phi(u_\star)\right\}.$$

Thus $\mathcal{V}_{\text{frontier}}$ consists of samples that lie in the model generable frontier, and are aligned with a valid direction of expansion. While $\mathcal{V}_{\text{frontier}}$ does not provide a general first-principles characterization of all useful samples, it formally captures in a minimal way the qualitative tension we want to study: useful self-generated samples should plausibly be (i) *generable* by the current model, yet (ii) *not* lying within its dominant modes, and (iii) aligned with valid directions of expansion. The next result shows that standard self-generation scheme might sample data within this natural class of frontier-valid samples with extremely low probability even in the introduced illustrative Gaussian generator setting.

Proposition 1 (Standard self-generation is unlikely to find frontier-valid samples)**.** *Let $X \sim p_0$, and let $\mathcal{V}_{\text{frontier}}$ be defined above. For any $\eta \in (0,1)$, define $\rho_\eta := \sigma\sqrt{m + 2\sqrt{m\log(1/\eta)} + 2\log(1/\eta)}$. If $\rho_- = \rho_\eta$, then*

$$\Pr\big(X \in \mathcal{V}_{\text{frontier}}\big) \le \eta \exp\left(-\frac{m-1}{2}\cos^2\phi\right).$$

Consequently, for N i.i.d. passive samples $X_1, \dots, X_N \sim p_0$,

$$\Pr\left(\exists i \in [N] : X_i \in \mathcal{V}_{\text{frontier}}\right) \le N\eta \exp\left(-\frac{m-1}{2}\cos^2\phi\right).$$

What the proposition reveals. The proposition isolates two sources of difficulty in recursive self-generation. The η-term reflects the rarity of reaching the generable frontier at all, namely of producing a sample with substantial orthogonal deviation from the dominant modes. Beyond this, an additional exponential penalty governs the chance of landing in the correct valid cone. Hence, when valid expansion directions are sparse, the sample complexity of passively obtaining a frontier-valid point becomes exponential in the orthogonal dimension m.

The algorithmic implication is clear: closed-loop verifier filtering alone is not enough for reliable expansion. Successful expansion requires actively steering generation toward the generable valid frontier.

B.2. Extensive Analysis with Proofs

Gaussian warm-up: self-generation requires rare valid frontier events We present a simple Gaussian model that isolates the two difficulties behind standard self-generation for efficient out-of-distribution discovery: sampling a *large deviation* away from the low-dimensional region captured by pre-training, and doing so along a *valid direction* of expansion.

We work directly on the design space $\mathcal{X} = \mathbb{R}^d$, so in this warm-up there is no separate representation map and $\mathcal{X} = \mathcal{Z}$. Let $U \subset \mathbb{R}^d$ be a k-dimensional linear subspace, and write

$$m := d - k, \qquad \mathbb{R}^d = U \oplus U^\perp.$$

For every $x \in \mathbb{R}^d$, denote by

$$x_U := \Pi_U x, \qquad x_\perp := \Pi_{U^\perp} x,$$

so that $x = x_U + x_\perp$. Throughout this subsection we assume $m \geq 2$, since the directional effect of interest only appears when the orthogonal complement has dimension at least two.

Pre-trained model. We model a pre-trained generator p^{θ_0} by the anisotropic Gaussian

$$X \sim p^{\theta_0} =: p_0 = \mathcal{N}\big(0,\, I_U \oplus \sigma^2 I_{U^\perp}\big), \qquad 0 < \sigma \ll 1.$$

Equivalently,

$$X_U \sim \mathcal{N}(0, I_k), \qquad X_\perp \sim \mathcal{N}(0, \sigma^2 I_m),$$

independently. Its density is

$$p_0(x) = (2\pi)^{-d/2}\sigma^{-m} \exp\left(-\frac{1}{2}\Big(\|x_U\|_2^2 + \sigma^{-2}\|x_\perp\|_2^2\Big)\right).$$

Generable set. Fix a level $\epsilon \in (0, (2\pi)^{-d/2}\sigma^{-m})$. In the sense of Definition 1, the ϵ-level generable set of the pre-trained model is

$$\Omega^\epsilon_{\theta_0} := \{x \in \mathbb{R}^d : p_0(x) \geq \epsilon\}.$$

For the Gaussian above this is the ellipsoid

$$\Omega^\epsilon_{\theta_0} = \left\{x \in \mathbb{R}^d : \|x_U\|_2^2 + \sigma^{-2}\|x_\perp\|_2^2 \leq r_\epsilon^2\right\},$$

where

$$r_\epsilon^2 := 2\log\left(\frac{(2\pi)^{-d/2}\sigma^{-m}}{\epsilon}\right).$$

Valid direction and frontier seed set. Fix a unit vector $u_\star \in U^\perp$ and an opening angle $\phi \in (0, \pi/2)$. We define the one-sided cone of valid orthogonal directions by

$$C_\phi(u_\star) := \left\{y \in U^\perp \setminus \{0\} : \left\langle \frac{y}{\|y\|_2}, u_\star \right\rangle \geq \cos\phi\right\}.$$

We also fix a radius $R_U > 0$ and shell radii $0 < \rho_- < \rho_+$ satisfying

$$R_U < r_\epsilon, \qquad \rho_+ < \sigma\sqrt{r_\epsilon^2 - R_U^2}.$$

The corresponding *valid frontier seed set* is

$$\mathcal{V}_{\text{seed}} := \left\{x \in \mathbb{R}^d : \|x_U\|_2 \leq R_U,\ \rho_- \leq \|x_\perp\|_2 \leq \rho_+,\ x_\perp \in C_\phi(u_\star)\right\}.$$

These are rare points that are still inside the current generable set, but already lie on a specific valid orthogonal frontier along which future continued pre-training may expand.

Lemma B.1 (Frontier seeds lie inside the current generable set)**.** *Under the above choice of R_U, ρ_+, ϵ, one has*

$$\mathcal{V}_{\text{seed}} \subseteq \Omega^\epsilon_{\theta_0}.$$

Proof. Take any $x \in \mathcal{V}_{\text{seed}}$. Then $\|x_U\|_2 \leq R_U$ and $\|x_\perp\|_2 \leq \rho_+$, hence

$$\|x_U\|_2^2 + \sigma^{-2}\|x_\perp\|_2^2 \leq R_U^2 + \sigma^{-2}\rho_+^2 < r_\epsilon^2.$$

By the explicit description of $\Omega^\epsilon_{\theta_0}$, this implies $x \in \Omega^\epsilon_{\theta_0}$. □

A larger valid design space. We consider a valid design space that extends beyond the current generable set in the same direction $u_\star$. Fix any $R > \sigma\sqrt{r_\epsilon^2 - R_U^2}$, and define

$$\mathcal{V}_{\text{out}} := \left\{ x \in \mathbb{R}^d : \|x_U\|_2 \le R_U,\ \sigma\sqrt{r_\epsilon^2 - R_U^2} < \|x_\perp\|_2 \le R,\ x_\perp \in C_\phi(u_\star) \right\}.$$

We then set

$$S_0 := \left\{ x \in \mathbb{R}^d : \|x_U\|_2 \le R_U,\ \|x_\perp\|_2 \le \rho_- \right\}, \qquad \Omega^\star := S_0 \cup \mathcal{V}_{\text{seed}} \cup \mathcal{V}_{\text{out}}.$$

Thus $\Omega^\star$ contains an already-valid core S_0, a rare but still generable frontier $\mathcal{V}_{\text{seed}}$, and a genuinely out-of-distribution valid region $\mathcal{V}_{\text{out}} \subset (\Omega^\epsilon_{\theta_0})^c$. The role of the negative result below is to show that, even before trying to reach $\mathcal{V}_{\text{out}}$, passive self-generation is already unlikely to find the expansion-enabling seeds $\mathcal{V}_{\text{seed}}$.

A convenient large-deviation scale. Since $X_\perp \sim \mathcal{N}(0, \sigma^2 I_m)$, one has

$$\frac{\|X_\perp\|_2^2}{\sigma^2} \sim \chi^2_m.$$

Therefore, for any $\eta \in (0,1)$, the choice

$$\rho_\eta := \sigma\sqrt{m + 2\sqrt{m\log(1/\eta)} + 2\log(1/\eta)}$$

satisfies

$$\Pr\big(\|X_\perp\|_2 \ge \rho_\eta\big) \le \eta$$

by the Laurent-Massart inequality (Laurent & Massart, 2000; Boucheron et al., 2013).

Proposition 1 (Standard self-generation is unlikely to find frontier-valid samples)**.** *Let $X \sim p_0$, and let $\mathcal{V}_{\text{frontier}}$ be defined above. For any $\eta \in (0,1)$, define $\rho_\eta := \sigma\sqrt{m + 2\sqrt{m\log(1/\eta)} + 2\log(1/\eta)}$. If $\rho_- = \rho_\eta$, then*

$$\Pr\big(X \in \mathcal{V}_{\text{frontier}}\big) \le \eta \exp\left(-\frac{m-1}{2}\cos^2\phi\right).$$

Consequently, for N i.i.d. passive samples $X_1, \dots, X_N \sim p_0$,

$$\Pr\left(\exists i \in [N] : X_i \in \mathcal{V}_{\text{frontier}}\right) \le N\eta \exp\left(-\frac{m-1}{2}\cos^2\phi\right).$$

Proof. By definition of $\mathcal{V}_{\text{seed}}$,

$$\{X \in \mathcal{V}_{\text{seed}}\} \subseteq \{\|X_\perp\|_2 \ge \rho_-\} \cap \left\{ \left\langle \frac{X_\perp}{\|X_\perp\|_2}, u_\star \right\rangle \ge \cos\phi \right\}.$$

Therefore

$$\Pr\big(X \in \mathcal{V}_{\text{seed}}\big) \le \Pr\left(\|X_\perp\|_2 \ge \rho_-, \left\langle \frac{X_\perp}{\|X_\perp\|_2}, u_\star \right\rangle \ge \cos\phi \right).$$

Now write

$$X_\perp = \sigma G, \qquad G \sim \mathcal{N}(0, I_m).$$

Let

$$R := \|G\|_2, \qquad S := \frac{G}{\|G\|_2} \in \mathbb{S}^{m-1}.$$

For an isotropic Gaussian, R and S are independent, and S is uniform on the unit sphere $\mathbb{S}^{m-1}$. Since $X_\perp = \sigma G$, the previous probability equals

$$\Pr\left(R \ge \frac{\rho_-}{\sigma}, \langle S, u_\star\rangle \ge \cos\phi\right) = \Pr\left(R \ge \frac{\rho_-}{\sigma}\right) \cdot \Pr\left(\langle S, u_\star\rangle \ge \cos\phi\right),$$

which proves

$$\Pr\big(X \in \mathcal{V}_{\text{seed}}\big) \le \Pr\big(\|X_\perp\|_2 \ge \rho_-\big) \cdot \Pr\left(\left\langle \frac{X_\perp}{\|X_\perp\|_2}, u_\star \right\rangle \ge \cos\phi\right).$$

For the directional term, a standard spherical-cap bound for $S \sim \mathrm{Unif}(\mathbb{S}^{m-1})$ gives, for every $t \in [0,1]$,

$$\Pr\big(\langle S, u_\star\rangle \ge t\big) \le \exp\left(-\frac{m-1}{2}t^2\right).$$

Applying this with $t = \cos\phi$ yields

$$\Pr\left(\left\langle \frac{X_\perp}{\|X_\perp\|_2}, u_\star \right\rangle \ge \cos\phi\right) = \Pr\big(\langle S, u_\star\rangle \ge \cos\phi\big) \le \exp\left(-\frac{m-1}{2}\cos^2\phi\right).$$

Combining the two bounds proves

$$\Pr\big(X \in \mathcal{V}_{\text{seed}}\big) \le \Pr\big(\|X_\perp\|_2 \ge \rho_-\big) \exp\left(-\frac{m-1}{2}\cos^2\phi\right).$$

If $\rho_- = \rho_\eta$, then by the above χ^2-tail bound,

$$\Pr\big(\|X_\perp\|_2 \ge \rho_\eta\big) \le \eta,$$

hence

$$\Pr\big(X \in \mathcal{V}_{\text{seed}}\big) \le \eta \exp\left(-\frac{m-1}{2}\cos^2\phi\right).$$

Finally, for N i.i.d. passive samples $X_1, \dots, X_N$, the union bound gives

$$\Pr(\exists i \in [N] : X_i \in \mathcal{V}_{\text{seed}}) \le \sum_{i=1}^{N} \Pr(X_i \in \mathcal{V}_{\text{seed}}) = N\ \Pr(X \in \mathcal{V}_{\text{seed}}) \le N\eta \exp\left(-\frac{m-1}{2}\cos^2\phi\right).$$

□

C. Gradient Descent and Ascent on flow matching loss

C.1. Signed replay-based continued pre-training and connection to data unlearning

This appendix formalizes the signed continued-pretraining update used in Sec. 4 and clarifies its connection to data unlearning objectives for diffusion models (Alberti et al., 2025). The key point is that, after each active-query round, verifier feedback naturally partitions the queried synthetic data into an *accepted* set and a *rejected* set. This induces a retain/forget decomposition analogous to data unlearning, except that here the partition is generated online by the verifier and used to improve out-of-distribution validity coverage.

Per-round replay buffers. At round t, let

$$\mathcal{D}_t^+ \subseteq \{x_i : y_i = 1,\ i \le t\}, \qquad \mathcal{D}_t^- \subseteq \{x_i : y_i = 0,\ i \le t\}$$

denote the accepted and rejected replay buffers accumulated up to round t. Thus the queried synthetic data available for continued pre-training are partitioned into

$$X_t := \mathcal{D}_t^+ \uplus \mathcal{D}_t^-, \qquad A_t := \mathcal{D}_t^-,$$

where $\uplus$ denotes disjoint union of indexed samples. By construction,

$$A_t \subseteq X_t, \qquad X_t \setminus A_t = \mathcal{D}_t^+.$$

This is exactly the retain/forget decomposition considered in data unlearning: X_t is the current empirical dataset, A_t is the subset to be unlearned, and $X_t \setminus A_t$ is the retained subset.

Underlying flow-matching loss. Let $\ell(\theta; x)$ denote the standard flow-matching training loss on a sample $x \in \mathcal{X}$, namely the same loss used in the original pre-training objective. For any finite empirical collection S, define

$$L_S(\theta) := \frac{1}{|S|} \sum_{x \in S} \ell(\theta; x). \tag{23}$$

In particular,

$$L_{X_t \setminus A_t}(\theta) = L_{\mathcal{D}_t^+}(\theta), \qquad L_{A_t}(\theta) = L_{\mathcal{D}_t^-}(\theta).$$

Reduction to the deletion objective. A first natural objective is to train the model as if the rejected samples had been removed from the replay set, namely

$$L_t^{\mathrm{del}}(\theta) := L_{X_t \setminus A_t}(\theta) = L_{\mathcal{D}_t^+}(\theta). \tag{24}$$

Writing

$$m_t := |\mathcal{D}_t^+|, \qquad k_t := |\mathcal{D}_t^-|, \qquad n_t := m_t + k_t,$$

the same algebra used in (Alberti et al., 2025) gives the exact decomposition

$$L_t^{\mathrm{del}}(\theta) = \frac{n_t}{m_t} L_{X_t}(\theta) - \frac{k_t}{m_t} L_{A_t}(\theta) = \frac{n_t}{m_t} L_{X_t}(\theta) - \frac{k_t}{m_t} L_{\mathcal{D}_t^-}(\theta). \tag{25}$$

Thus, even the valid-only objective can be rewritten using *both* accepted and rejected samples. This is precisely the same deletion identity underlying SISS in (Alberti et al., 2025), specialized here to the verifier-induced partition of queried synthetic data. In the terminology of (Alberti et al., 2025), our practical implementation uses the corresponding *non-importance-sampled* variant, i.e., the analogue of SISS (No IS), rather than the one-pass importance-sampled estimator.

Why we use a signed objective. For out-of-distribution generable-set expansion, Eq. (24) is often too conservative: it treats rejected samples only as points to be discarded. In contrast, rejected verifier queries contain useful information about directions in design space toward which the current model should allocate *less* probability mass. This motivates the signed replay objective used in the main text:

$$L_t^{\mathrm{signed}}(\theta) := L_{\mathcal{D}_t^+}(\theta) - \alpha_t\, L_{\mathcal{D}_t^-}(\theta). \tag{26}$$

Its gradient is exactly the signed update:

$$\nabla_\theta L_t^{\text{signed}}(\theta_t) = \nabla_\theta L_{\mathcal{D}_t^+}(\theta_t) - \alpha_t \nabla_\theta L_{\mathcal{D}_t^-}(\theta_t). \tag{27}$$

At the minibatch level, our implementation follows the *non-importance-sampled* variant: we draw separate minibatches

$$U_t^+ \subseteq \mathcal{D}_t^+, \qquad U_t^- \subseteq \mathcal{D}_t^-,$$

and form the empirical losses

$$\widehat{L}_t^+(\theta) = \frac{1}{|U_t^+|} \sum_{x \in U_t^+} \ell(\theta; x), \qquad \widehat{L}_t^-(\theta) = \frac{1}{|U_t^-|} \sum_{x \in U_t^-} \ell(\theta; x).$$

The practical signed update is then

$$g_t = \nabla \widehat{L}_t^+(\theta_t) - \alpha_t \nabla \widehat{L}_t^-(\theta_t).$$

Hence accepted samples provide an attractive update, while rejected samples provide a repulsive update of controlled magnitude. This corresponds to the analogue of SISS (No IS) in (Alberti et al., 2025): two separate replay-buffer minibatches are used, rather than a single importance-sampled mixture minibatch.

Relation to weighted data-unlearning objectives. The weighted SISS objective of (Alberti et al., 2025) can be written as

$$L^{\text{wSISS}}(\theta) = L_{X \setminus A}(\theta) - s \frac{|A|}{|X \setminus A|} L_A(\theta). \tag{28}$$

Applying this to the verifier-induced partition

$$X = X_t = \mathcal{D}_t^+ \uplus \mathcal{D}_t^-, \qquad A = A_t = \mathcal{D}_t^-,$$

yields

$$L_t^{\text{wSISS}}(\theta) = L_{\mathcal{D}_t^+}(\theta) - s_t \frac{k_t}{m_t} L_{\mathcal{D}_t^-}(\theta). \tag{29}$$

Therefore our signed replay objective in Eq. (26) is exactly the same class of weighted retain-minus-forget objective, with the identification

$$\alpha_t = s_t \frac{k_t}{m_t}. \tag{30}$$

In this sense, the update used by ACTFLOW is a direct reduction of the weighted unlearning objective of (Alberti et al., 2025) to the online verifier-guided expansion setting considered here.

Possible importance sampling (IS) extension. As mentioned, our implementation is the direct analogue of SISS (No IS) in (Alberti et al., 2025), rather than the one-pass importance-sampled estimator introduced there. An importance-sampled extension could also be used in our setting by sampling from a suitable mixture over accepted and rejected replay samples and reweighting by the corresponding importance ratios, but we avoid this additional estimator complexity here for simplicity.

Interpretation in our setting. The reduction above is exact at the level of the empirical replay buffers held fixed at round t. Conditionally on the current buffers $(\mathcal{D}_t^+, \mathcal{D}_t^-)$, the signed objective

$$L_t^{\text{signed}}(\theta) = L_{\mathcal{D}_t^+}(\theta) - \alpha_t L_{\mathcal{D}_t^-}(\theta)$$

should be read as follows:

1. the first term increases likelihood of synthetic samples that have been certified as valid by the verifier and therefore support expansion into new valid regions;
2. the second term decreases likelihood of queried samples that were explicitly rejected by the verifier and therefore encode evidence against those directions of expansion.

Thus, unlike standard valid-only pre-training, the signed replay update uses *both* outcomes of the verifier query and converts the online generation-and-verification loop into a principled retain/forget training signal.

Practical choice of α_t. In the practical implementation, the loss-level weight α_t is calibrated online as a gradient-norm fraction. Concretely, for a target ratio $\alpha \in [0, 1]$, we set

$$\alpha_t \;:=\; \alpha \, \frac{\|\nabla \widehat{L}_t^{+}(\theta_t)\|_2}{\|\nabla \widehat{L}_t^{-}(\theta_t)\|_2},$$

This yields

$$\left\|\alpha_t \nabla \widehat{L}_t^{-}(\theta_t)\right\|_2 \approx \alpha \left\|\nabla \widehat{L}_t^{+}(\theta_t)\right\|_2,$$

so the rejected-sample contribution is scaled to have a prescribed norm fraction relative to the accepted-sample contribution.

D. Intermezzo: Theory of Active Safe Logistic Regression

Suppose that given any query point x, the verifier provides binary feedback $y \in \{0, 1\}$ generated according to

$$\Pr[y = 1 \mid x] = s(f(x)), \tag{31}$$

where $s : \mathbb{R} \to [0, 1]$ is the sigmoid function, and $f : \mathcal{X} \to \mathbb{R}$ is a validity function. We assume that f lies in the RKHS $\mathcal{H}_k$ associated with a kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, with bounded norm $\|f\|_k \le B$. We also assume that f is L_f-smooth with respect to a metric d, namely

$$|f(x) - f(x')| \le L_f d(x, x') \qquad \forall x, x' \in \mathcal{X}.$$

Without loss of generality, we further assume that $k(x, x') \le 1$ for all $x, x' \in \mathcal{X}$.

We adopt a *pre-query* indexing convention. At the beginning of round t, the learner has access to the history

$$H_{t-1} \triangleq \{(x_\tau, y_\tau)\}_{\tau=1}^{t-1},$$

fits a probabilistic model using H_{t-1}, constructs a safe set S_t, and then selects the next query x_t. Given H_{t-1}, the learner estimates f by minimizing the regularized negative log-likelihood

$$\begin{aligned} \mu_t &\triangleq \operatorname*{arg\,min}_{g \in \mathcal{H}_k,\ \|g\|_k \le B} \mathcal{L}(g, H_{t-1}), \\ \mathcal{L}(g, H_{t-1}) &\triangleq \sum_{\tau=1}^{t-1} \Big[-y_\tau \log\big(s(g(x_\tau))\big) - (1 - y_\tau) \log\big(1 - s(g(x_\tau))\big) \Big] + \frac{\lambda}{2} \|g\|_k^2 , \end{aligned} \tag{32}$$

where $\lambda > 0$ is a regularization coefficient. By the Representer Theorem (Schölkopf et al., 2001), the solution lies in the span of the kernel sections at the previously queried points.

Given μ_t, the learner predicts the verifier output via $s(\mu_t(x))$. Previous work gives anytime-valid confidence sets of the form

$$[s(\mu_t(x)) \pm \beta_t(\delta)\sigma_t(x)]$$

for kernelized logistic regression (Pásztor et al., 2024). Under our indexing convention, the uncertainty scores are

$$\sigma_t^2(x) = k(x, x) - k_{t-1}^\top(x) \left(K_{t-1} + \lambda\kappa I_{t-1}\right)^{-1} k_{t-1}(x),$$

where

$$\kappa \triangleq \sup_{a \le B} \frac{1}{\dot{s}(a)},$$

$k_{t-1}(x)$ is the kernel vector with entries $[k_{t-1}(x)]_i = k(x_i, x)$, and K_{t-1} is the kernel matrix with entries $[K_{t-1}]_{i,j} = k(x_i, x_j)$.

We will use the following pre-query reindexing of the confidence-sequence guarantee.

Theorem D.1 (Kernelized Logistic Confidence Sequences)**.** *Assume $f \in \mathcal{H}_k$ and $\|f\|_k \le B$. Assume that the data $x_1, \dots, x_t$ used to fit the model μ_t lie in a compact subset $A \subseteq \mathcal{X}$. Let $0 < \delta < 1$ and define*

$$\beta_t(\delta) \triangleq 4L_s B + 2L_s \sqrt{\frac{2\kappa}{\lambda}\Big(\gamma_{t-1}^A + \log(1/\delta)\Big)}, \tag{33}$$

where

$$\gamma_t^A \triangleq \max_{x_1, \dots, x_t \in A} \frac{1}{2} \log\det\Big(I_t + (\lambda\kappa)^{-1} K_t\Big), \qquad L_s \triangleq \sup_{a \le B} \dot{s}(a).$$

Then

$$\Pr\Big(\forall t \ge 1,\ \forall x \in A :\ \big|s(\mu_t(x)) - s(f(x))\big| \le \beta_t(\delta)\, \sigma_t(x)\Big) \ge 1 - \delta.$$

Using these confidence sets, the learner may perform active safe logistic regression, querying only points that are certified to satisfy $s(f(x)) \ge h$ for some threshold $h \in [0, 1]$. This parallels safe Bayesian optimization in the regression setting (e.g., Sui et al., 2015; 2018).

In this setting, safe designs are those for which the verifier returns label 1 with sufficiently high probability. Suppose the learner is given a safe seed set $S_0 \subseteq \mathcal{X}$. At round t, the learner updates its certified safe set as

$$S_t = S_{t-1} \cup \{x \in \mathcal{X} \mid \exists x' \in S_{t-1} : s(\mu_t(x')) - \beta_t(\delta)\sigma_t(x') - L_s L_f d(x, x') \geq h\}. \tag{34}$$

The next lemma shows that, conditioned on the confidence event of Theorem D.1, every point ever added to S_t is indeed safe.

Lemma D.2 (Safety)**.** *Condition on the event that the confidence intervals of Theorem D.1 are valid. Then for every $t \geq 0$, every $x \in S_t$ satisfies $s(f(x)) \geq h$.*

Proof. We argue by induction on t. The claim holds at $t = 0$ by assumption on the seed set S_0. Now fix $t \geq 1$, and let $x \in S_t$. Either $x \in S_{t-1}$, in which case the claim follows by induction, or

$$x \in \{x \in \mathcal{X} \mid \exists x' \in S_{t-1} : s(\mu_t(x')) - \beta_t(\delta)\sigma_t(x') - L_s L_f d(x, x') \geq h\}.$$

In the latter case, there exists $x' \in S_{t-1}$ such that

$$s(\mu_t(x')) - \beta_t(\delta)\sigma_t(x') - L_s L_f d(x, x') \geq h.$$

By the confidence event,

$$s(f(x')) \geq s(\mu_t(x')) - \beta_t(\delta)\sigma_t(x').$$

Moreover, since s is L_s-Lipschitz and f is L_f-Lipschitz,

$$s(f(x)) \geq s(f(x')) - L_s L_f d(x, x').$$

Combining the two inequalities gives

$$s(f(x)) \geq s(\mu_t(x')) - \beta_t(\delta)\sigma_t(x') - L_s L_f d(x, x') \geq h.$$

Thus every $x \in S_t$ is safe. □

Notice that by construction, the sequence $\{S_t\}_{t \geq 0}$ is monotone:

$$S_{t-1} \subseteq S_t \qquad \forall t \geq 1.$$

Using this set of safe decisions, the learner may follow an active learning procedure which samples the next point at which to query the verifier as the one in the set of safe decisions that has the highest uncertainty score, often referred to as *safe uncertainty sampling*:

$$x_t = \arg\max_{x \in S_t} \sigma_t(x). \tag{35}$$

After querying the verifier at x_t and observing y_t, the history updates as

$$H_t = H_{t-1} \cup \{(x_t, y_t)\}.$$

We ultimately want to show some notion of suitable expansion of the safe set by following the above sampling procedure. As our safety set at any time is defined as an expansion of the safe set at the previous time, we cannot guarantee convergence to the entire subset of $\mathcal{X}$ which is safe. Instead, we must define some notion of *reachable safe set* of decisions. To this end, we consider the tightened one step reachability operator defined as

$$R_\varepsilon(S) \triangleq \{x \in \mathcal{X} \mid \exists x' \in S : s(f(x')) - L_s L_f d(x, x') - \varepsilon \geq h\}.$$

The H-fold recursive application of this operator is denoted by $R_\varepsilon^H(S)$, and its closure as $H \to \infty$ is denoted $\bar{R}_\varepsilon(S)$. Our goal is to show that safe uncertainty sampling expands the certified safe set toward $\bar{R}_0(S_0)$. We first prove an uncertainty-reduction bound over the current safe set. Let

$$\Omega^\star \triangleq \{x \in \mathcal{X} : s(f(x)) \geq h\}$$

denote the subset of $\mathcal{X}$ corresponding to the *valid design space*, and assume that $\Omega^\star$ is compact.

Lemma D.3 (Uncertainty Reduction). *Fix $t \geq 0$ and $T \geq 1$. Under safe uncertainty sampling (35), the epistemic uncertainty at time $t+T$ decays as*

$$\max_{x \in S_t} \sigma_{t+T}(x) \leq \sqrt{\frac{2\bar{\sigma}^2}{\log(1+(\lambda\kappa)^{-1}\bar{\sigma}^2)}} \sqrt{\frac{\gamma_T^{\Omega^\star}}{T}},$$

where $\bar{\sigma} \triangleq \max_{x \in \mathcal{X}} \sigma_1(x)$.

Proof. By monotonicity of the uncertainty score, it holds that

$$\begin{aligned}
\max_{x \in S_t} \sigma_{t+T}(x) &\leq \frac{1}{T} \sum_{n=0}^{T-1} \max_{x \in S_t} \sigma_{t+n}(x) && \text{(uncertainty is monotone)} \\
&\leq \frac{1}{T} \sum_{n=0}^{T-1} \max_{x \in S_{t+n}} \sigma_{t+n}(x) && (S_t \text{ are monotone}) \\
&= \frac{1}{T} \sum_{n=0}^{T-1} \sigma_{t+n}(x_{t+n}) && \text{(safe uncertainty sampling (35))} \\
&\leq \frac{1}{\sqrt{T}} \sqrt{\sum_{n=0}^{T-1} \sigma_{t+n}^2(x_{t+n})} && \text{(Cauchy–Schwarz).}
\end{aligned}$$

It holds that $x \leq \frac{\bar{x}}{\log(1+\bar{x})} \log(1+x)$ for $0 \leq x \leq \bar{x}$. Let $K_{t:t+T-1}$ be the Gram matrix of the queried points $x_t, \ldots, x_{t+T-1}$, let $K_{1:t-1}$ be the Gram matrix of the previously queried points $x_1, \ldots, x_{t-1}$, let $K_{t:t+T-1,\,1:t-1}$ be the corresponding cross-Gram matrix, and define

$$\Sigma_{t:t+T-1|1:t-1} \triangleq K_{t:t+T-1} - K_{t:t+T-1,\,1:t-1}(K_{1:t-1} + \lambda\kappa I_{t-1})^{-1} K_{1:t-1,\,t:t+T-1}.$$

Then the above inequality implies that

$$\begin{aligned}
\max_{x \in S_t} \sigma_{t+T}(x) &\leq \frac{1}{\sqrt{T}} \sqrt{\frac{\bar{\sigma}^2}{\log(1+(\lambda\kappa)^{-1}\bar{\sigma}^2)}} \sqrt{\sum_{n=0}^{T-1} \log\left(1 + \frac{\sigma_{t+n}^2(x_{t+n})}{\lambda\kappa}\right)} \\
&= \frac{1}{\sqrt{T}} \sqrt{\frac{\bar{\sigma}^2}{\log(1+(\lambda\kappa)^{-1}\bar{\sigma}^2)}} \sqrt{\log\det(I_T + (\lambda\kappa)^{-1}\Sigma_{t:t+T-1|1:t-1})} \\
&\leq \frac{1}{\sqrt{T}} \sqrt{\frac{\bar{\sigma}^2}{\log(1+(\lambda\kappa)^{-1}\bar{\sigma}^2)}} \sqrt{\log\det(I_T + (\lambda\kappa)^{-1}K_{t:t+T-1})} \\
&\leq \frac{1}{\sqrt{T}} \sqrt{\frac{\bar{\sigma}^2}{\log(1+(\lambda\kappa)^{-1}\bar{\sigma}^2)}} \sqrt{\max_{x_1,\ldots,x_T \in \Omega^\star} \log\det(I_T + (\lambda\kappa)^{-1}K_T)} \\
&= \sqrt{2}\frac{\sqrt{\gamma_T^{\Omega^\star}}}{\sqrt{T}} \sqrt{\frac{\bar{\sigma}^2}{\log(1+(\lambda\kappa)^{-1}\bar{\sigma}^2)}}.
\end{aligned}$$

□

The next lemma shows that once the uncertainty over the current safe set is sufficiently small, additional safe uncertainty sampling expands the safe set according to the one-step reachability operator.

Lemma D.4 (One-step Reachable Expansion). *Consider $\varepsilon > 0$, $t \geq 0$, and $T \geq 1$. Suppose that*

$$T \geq \frac{8\beta_{t+T}^2 \gamma_T^{\Omega^\star}(\lambda\kappa + \bar{\sigma}^2)}{\varepsilon^2}.$$

Then, conditioned on the event that the intervals of Theorem D.1 are valid, it holds that

$$R_\varepsilon(S_t) \subseteq S_{t+T}.$$

Proof. By Lemma D.3 and the condition on T,

$$\max_{x \in S_t} \sigma_{t+T}(x) \leq \frac{\varepsilon}{2\beta_{t+T}}. \tag{36}$$

Consider any point $x \in R_\varepsilon(S_t)$. By definition, there exists $x' \in S_t$ such that

$$h \leq s(f(x')) - \varepsilon - L_s L_f d(x, x').$$

By the confidence event at round $t + T$,

$$s(f(x')) \leq s(\mu_{t+T}(x')) + \beta_{t+T}\sigma_{t+T}(x').$$

Hence

$$\begin{aligned} h &\leq s(f(x')) - \varepsilon - L_s L_f d(x, x') \\ &\leq s(\mu_{t+T}(x')) + \beta_{t+T}\sigma_{t+T}(x') - \varepsilon - L_s L_f d(x, x') \\ &\leq s(\mu_{t+T}(x')) - \beta_{t+T}\sigma_{t+T}(x') - L_s L_f d(x, x'), \end{aligned}$$

where the last step uses (36), since $x' \in S_t$ implies

$$\beta_{t+T}\sigma_{t+T}(x') \leq \frac{\varepsilon}{2}.$$

By monotonicity of the safe sets, $x' \in S_t \subseteq S_{t+T-1}$. Therefore there exists $x' \in S_{t+T-1}$ such that

$$s(\mu_{t+T}(x')) - \beta_{t+T}\sigma_{t+T}(x') - L_s L_f d(x, x') \geq h.$$

By the definition of S_{t+T} in (34), this implies $x \in S_{t+T}$. □

Applying the above result recursively leads to H-step expansion of S_0 at time $T^\star$. In particular, suppose $T^\star = HT$ for some T satisfying

$$T \geq \frac{8\beta_{HT}^2 \gamma_T^{\Omega^\star}(\lambda\kappa + \bar{\sigma}^2)}{\varepsilon^2}.$$

This implies that the condition of the above lemma holds for each interval of length T, since β_t is monotonically increasing. Consequently, it holds that

$$R_\varepsilon^H(S_0) \subseteq S_{T^\star}.$$

We can find $T^\star$ large enough to satisfy this condition as long as the complexity term $\beta_t^2 \gamma_t^{\Omega^\star}$ grows sublinearly.

E. Theoretical Analysis of Diffusion Models Active Diffusion Expansion

In the following, we first (i) report the general analysis for safe logistic regression that we introduced and presented in Apx. D, where we now re-interpret safety as validity, then we (ii) introduce a probabilistic modeling framework of the generative density p_t via energy-based models, and (iii) derive coverage guarantees with sample complexity for the proposed active expansion algorithm.

E.1. Probabilistic Modeling of Binary Verifier over Generative Model Learned Representation

We denote by $\phi : \mathcal{X} \to \mathcal{Z}$ the representation map learned by the pre-trained generative model. We now describe the active learning process of the verifier over the learned representation space $\mathcal{Z}$. Suppose that given any query point x with latent representation $z := \phi(x)$, the verifier provides binary feedback $y \in \{0, 1\}$ generated according to

$$\Pr[y = 1 \mid z] = s(g(z)), \tag{37}$$

where $s : \mathbb{R} \to [0, 1]$ is the sigmoid function, and $g : \mathcal{Z} \to \mathbb{R}$ is an unknown validity function. We assume that g lies in the RKHS $\mathcal{H}_k$ associated with a kernel $k : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$, with bounded norm $\|g\|_k \le B$ and that g is L_g-smooth with respect to a metric d, namely

$$|g(z) - g(z')| \le L_g d(z, z') \qquad \forall z, z' \in \mathcal{Z},$$

for B and L_g known. We additionally assume there exists a known constant $\bar{Z}$ such that $\int_{\mathcal{Z}} \exp(g(z))dz \le \bar{Z}$.

On unbounded domains with respect to Lebesgue measure, the energy condition is incompatible with globally bounded kernels such as the squared-exponential kernel. Indeed, if k is bounded on the diagonal and $\|g\|_k \le B$, then

$$|g(z)| \le \|g\|_k \sqrt{k(z, z)} \text{ for all } z \in \mathcal{Z}$$

implies that g is uniformly bounded. Hence $\exp(g)$ is bounded below by a positive constant, and therefore cannot be integrable over an infinite-measure domain. Thus, in this setting, the energy condition should be understood as requiring a kernel class capable of representing functions with sufficiently negative tails. For example, sums of bounded kernels with even-degree polynomial kernels yield RKHS that contain coercive negative polynomials, such as $g(z) = c - z^\top Az$ with $A \succ 0$, which satisfy the energy condition.

We adopt a *pre-query* indexing convention. At the beginning of round t, the learner has access to the history

$$H_{t-1} \triangleq \{(z_\tau, y_\tau)\}_{\tau=1}^{t-1}.$$

Given H_{t-1}, the learner estimates g by minimizing the regularized negative log-likelihood

$$\begin{aligned} \mu_t &\triangleq \operatorname*{arg\,min}_{u \in \mathcal{H}_k,\ \|u\|_k \le B} \mathcal{L}(u, H_{t-1}), \\ \mathcal{L}(u, H_{t-1}) &\triangleq \sum_{\tau=1}^{t-1} \Big[-y_\tau \log\big(s(u(z_\tau))\big) - (1 - y_\tau) \log\big(1 - s(u(z_\tau))\big) \Big] + \frac{\lambda}{2} \|u\|_k^2 , \end{aligned} \tag{38}$$

where $\lambda > 0$ is a regularization coefficient. By the Representer Theorem (Schölkopf et al., 2001), the solution lies in the span of the kernel sections at the previously queried points.

Given μ_t, the learner predicts the verifier output via $s(\mu_t(z))$. Previous work gives anytime-valid confidence sets of the form

$$[s(\mu_t(z)) \pm \beta_t \sigma_t(z)]$$

for kernelized logistic regression (Pásztor et al., 2024). Under our indexing convention, the uncertainty scores are

$$\sigma_t^2(z) = k(z, z) - k_{t-1}^\top(z) \left(K_{t-1} + \lambda\kappa I_{t-1}\right)^{-1} k_{t-1}(z),$$

where

$$\kappa \triangleq \sup_{a \le B} \frac{1}{\dot{s}(a)},$$

$k_{t-1}(z)$ is the kernel vector with entries $[k_{t-1}(z)]_i = k(z_i, z)$, and K_{t-1} is the kernel matrix with entries $[K_{t-1}]_{i,j} = k(z_i, z_j)$.

We will use the following pre-query reindexing of the confidence-sequence guarantee.

Theorem E.1 (Kernelized Logistic Confidence Sequences (Pásztor et al., 2024)). *Assume $g \in \mathcal{H}_k$ and $\|g\|_k \le B$. Assume that the queried points lie in a compact subset $A \subseteq \mathcal{Z}$. Without loss of generalization, suppose that $k(z, z') \le 1 \, \forall z, z' \in A$. Let $0 < \delta < 1$ and define*

$$\beta_t^A(\delta) \triangleq 4L_s B + 2L_s \sqrt{\frac{2\kappa}{\lambda}\Big(\gamma_{t-1}^A + \log(1/\delta)\Big)}, \tag{39}$$

where

$$\gamma_t^A \triangleq \max_{z_1,\dots,z_t \in A} \frac{1}{2} \log\det\Big(I_t + (\lambda\kappa)^{-1} K_t\Big), \qquad L_s \triangleq \sup_{a \le B} \dot{s}(a).$$

Then

$$\Pr\left(\forall t \ge 1,\ \forall z \in A:\ \big|s(\mu_t(z)) - s(g(z))\big| \le \beta_t^A(\delta)\, \sigma_t(z)\right) \ge 1 - \delta.$$

Using these confidence sets, the learner may perform guarded logistic regression, querying only points that are certified to satisfy $s(g(z)) \ge h$ for some threshold $h \in [0, 1]$. This parallels safe Bayesian optimization in the regression setting (e.g., Sui et al., 2015; 2018).

In this setting, safe designs are those for which the probabilistic verifier returns label 1 with sufficiently high probability. Suppose the learner is given a safe seed set $S_0 \subseteq \mathcal{Z}$. Given a monotonically increasing sequence of calibration coefficients at level δ, $\left\{\hat{\beta}_t(\delta)\right\}_{t \ge 0}$, the set of points that the learner can verify as safe may be expanded at round t as

$$S_t = S_{t-1} \cup \left\{ z \in \mathcal{Z} \;\middle|\; \exists z' \in S_{t-1} : s(\mu_t(z')) - \hat{\beta}_t(\delta)\sigma_t(z') - L_s L_g d(z, z') \ge h \right\}. \tag{40}$$

We differentiate between the sequence $\left\{\hat{\beta}_t(\delta)\right\}$ and $\left\{\beta_t^A(\delta)\right\}$ as computing the latter requires knowledge of the subset A, whereas our downstream analysis shows that our sampling remains restricted to an unknown set. Despite this, the learner may reasonably have an upper bound on the information capacity of the kernel restricted to the unknown subset, and thus may reasonably have access to such upper bounds.

The next lemma shows that, conditioned on the confidence event of Theorem E.1, every point ever added to S_t satisfies the validity condition.

Lemma E.2 (Safety). *Condition on the event that the confidence intervals of Theorem E.1 are valid at level $\delta \in (0, 1)$ under a compact set A with $k(z, z') \le 1 \, \forall z, z' \in A$ and suppose that the sequence $\left\{\hat{\beta}_t(\delta)\right\}_{t \ge 0}$ satisfies $\hat{\beta}_t(\delta) \ge \beta_t^A(\delta)$. Then for every $t \ge 0$, every $z \in S_t$ satisfies $s(g(z)) \ge h$.*

Proof. We argue by induction on t. The claim holds at $t = 0$ by assumption on the seed set S_0.

Now fix $t \ge 1$, and let $z \in S_t$. Either $z \in S_{t-1}$, in which case the claim follows by induction, or

$$z \in \left\{ z \in \mathcal{Z} \;\middle|\; \exists z' \in S_{t-1} : s(\mu_t(z')) - \hat{\beta}_t(\delta)\sigma_t(z') - L_s L_g d(z, z') \ge h \right\}.$$

In the latter case, there exists $z' \in S_{t-1}$ such that

$$s(\mu_t(z')) - \hat{\beta}_t(\delta)\sigma_t(z') - L_s L_g d(z, z') \ge h.$$

By the confidence event,

$$s(g(z')) \ge s(\mu_t(z')) - \hat{\beta}_t(\delta)\sigma_t(z').$$

Moreover, since s is L_s-Lipschitz and g is L_g-Lipschitz,

$$s(g(z)) \ge s(g(z')) - L_s L_g d(z, z').$$

Combining the two inequalities gives

$$s(g(z)) \ge s(\mu_t(z')) - \hat{\beta}_t(\delta)\sigma_t(z') - L_s L_g d(z, z') \ge h.$$

Thus every $z \in S_t$ exceeds the validity threshold. □

Notice that by construction, the sequence $\{S_t\}_{t \ge 0}$ is monotone:

$$S_{t-1} \subseteq S_t \qquad \forall t \ge 1.$$

E.2. Abstraction of the Generative Model as an EBM

Our primary abstraction of the generative model is that it approximately tracks the sequence of certified safe sets constructed with a sequence of calibration parameters $\left\{\hat{\beta}_t(\delta)\right\}$. We achieve this abstraction using an energy based model. Specifically, for a fixed $\gamma > 0$ and $0 < \ell \leq h$, we consider an energy function which satisfies

$$\begin{aligned} \mu_t' \in \Bigg\{ f : \int_{\mathcal{Z}} e^{f(z)} dz \leq \bar{Z}, \\ f(z) \geq \operatorname{logit}(h) \mathbf{1}(S_t) \, \forall z \in \mathcal{Z}, \\ f(z) \leq \operatorname{logit}(\ell) \text{ for all } z \text{ such that } d(z, S_t) \geq \gamma \Bigg\}. \end{aligned} \tag{41}$$

The lower bound on f captures the behavior where the updates to our generative model maintain high density on regions that our verifier certifies as valid according to (34). The upper bound on f instead captures the behavior where the model maintains low density on points far from this validated set. The explicit bound on the partition function $\bar{Z}$ ensures that the normalization constant does not cause the density placed on the valid region to vanish. These capture the desiderata that our generative model updates maintain large density in regions which have been verified as valid, and maintain small density in regions far away from this verified region. The generative model updates using both positive and negative samples, defined in Apx. C.1, are designed to satisfy these desiderata. For our theoretical analysis, we assume that this abstraction is valid.

Assumption E.1. *Consider $\gamma > 0$ and let ℓ be such that*

$$\frac{\exp(\operatorname{logit}(\ell))}{\underline{Z}} \leq \frac{\exp(\operatorname{logit}(h))}{\bar{Z}}. \tag{42}$$

where $\underline{Z} := \mathsf{vol}(S_0) \exp(\operatorname{logit}(h))$ and $\operatorname{logit}(u) := \log\left(\frac{u}{1-u}\right)$. We assume that the generative model can be abstracted as an EBM by considering the density

$$p_t(z) = \frac{\exp(\mu_t'(z))}{Z_t}, \qquad Z_t := \int_{z \in \mathcal{Z}} \exp(\mu_t'(z)) \, \mathrm{d}z. \tag{43}$$

for the energy function μ_t' defined in (41).

We define the corresponding *generable set* at level τ as the high-probability set

$$\Omega_t^\tau := \{x \in \mathcal{X} : p_t(z) \geq \tau\}. \tag{44}$$

The EBM-abstraction of the generative model allows us to draw a relationship between the generable set of the EBM and the certified safe sets in which the generable set at some level τ and round t contains the set S_t.

Lemma E.3 (Generable and Verifier Set Inequality for the Recursive Safe Set)**.** *Suppose that Assumption E.1 holds with some $\gamma > 0$.*

Then, for any

$$\tau \leq \frac{\exp(\operatorname{logit}(h))}{\bar{Z}}, \tag{45}$$

S_t is a subset of the high-probability set Ω_t^τ for the EBM model defined by Assumption E.1, namely $S_t \subseteq \Omega_t^\tau$.

Proof. By definition of the high-probability set,

$$\Omega_t^\tau = \{x \in \mathcal{X} : p_t^\pi(x) \geq \tau\} = \left\{x \in \mathcal{X} : \frac{\exp(\mu_t'(x))}{Z_t} \geq \tau\right\} = \{x \in \mathcal{X} : \mu_t'(x) \geq \log \tau + \log Z_t\}. \tag{46}$$

Now let $z \in S_t$. By definition of μ_t', $\mu_t'(z)) \geq \operatorname{logit}(h)$. Therefore, a sufficient condition for $x \in \Omega_t^\tau$ is $\operatorname{logit}(h) \geq \log \tau + \log Z_t$, or equivalently, $\tau \leq \exp(\operatorname{logit}(h))/Z_t$.

It remains to upper bound Z_t. This holds immediately by definition of μ'_t such that it satisfies

$$Z_t = \int_{\mathcal{Z}} \exp(\mu'_t(z))dz \leq \bar{Z}.$$

Plugging this bound in above, we find that a sufficient condition is,

$$\tau \leq \frac{\exp(\text{logit}(h))}{\bar{Z}}. \tag{47}$$

Under this condition, every $z \in S_t$ satisfies $\mu_t(z) \geq \log \tau + \log Z_t$, hence by (46) we have $z \in \Omega_t^\tau$. Therefore, $S_t \subseteq \Omega_t^\tau$, concluding the proof. □

At the same time, we can show that the generable set lies within an extension of the valid design space. To this end, we define the inflation of the valid design space.

Definition 3. *We define the inflation of the valid set by amount $\zeta \in \mathbb{R}$ in metric d as*

$$\Gamma(\zeta) = \{z \in \mathcal{Z} \setminus \Omega_* : d(z, \Omega_*) \leq \zeta\}.$$

Assumption E.2. *Let $\gamma > 0$ be the same as Assumption E.1. We suppose that the set $\Omega_* \cup \Gamma(\gamma)$ is compact. Further assume without loss of generality (by normalization of the kernel) that $k(z, z') \leq 1$ for all $z, z' \in \Omega_* \cup \Gamma(\gamma)$.*

Lemma E.4. *Fix some $\gamma > 0$. Suppose that the sequence of calibration coefficients $\left\{\hat{\beta}_t(\delta)\right\}$ used to construct the sets S_t satisfies $\hat{\beta}_t(\delta) \geq \beta_t^{\Omega_* \cup \Gamma(\gamma)}(\delta)$. Let Assumptions E.1-E.2 hold at level γ and condition on the calibration event of Theorem E.1 with $A \leftarrow \Omega_* \cup \Gamma(\gamma)$. For $\tau \geq \frac{\exp(\text{logit}(\ell))}{\underline{Z}}$ it holds that $\Omega_t^\tau \subseteq \Omega_* \cup \Gamma(\gamma)$.*

Proof. By (46) and the fact that $Z_t \geq \underline{Z}$ it holds that

$$\begin{aligned}
\Omega_t^\tau &\subseteq \{z \in \mathcal{Z} : \mu'_t(z) \geq \log(\tau) + \log(\underline{Z})\} \\
&\subseteq \{z \in \mathcal{Z} : \mu'_t(z) \geq \text{logit}(\ell)\} \\
&\subseteq S_t \cup \{z \in \mathcal{Z} \setminus S_t : d(z, S_t) \leq \gamma\} \\
&\subseteq \Omega_* \cup \{z \in \mathcal{Z} \setminus \Omega_* : d(z, \Omega_*) \leq \gamma\} \\
&= \Omega_* \cup \Gamma(\gamma),
\end{aligned}$$

where the final set inequality follows from Lemma E.2. □

E.3. Active Diffusion Expansion Core Analysis

By building on the two previous subsections, we can finally analyze the active expansion process of the set Ω_t^τ. Intuitively, we proceed in two steps: first, we analyze the expansion of the verifier valid set S_t under samples obtained via generative sampling; second, we use Lemma E.3 to transfer this expansion guarantee to the generable set Ω_t^τ, which is the object we ultimately wish to study.

We adopt the same pre-query indexing convention as in Section E.1. Accordingly, at round t the learner first constructs Ω_t^τ and the verifier model (μ_t, σ_t), and then selects the next query point. The sampling scheme used by Algorithm 1 is given by as follows.

Assumption E.3. *For some fixed $\alpha \geq 1$, we suppose that the sample points queried by the learner satisfy*

$$z_t \in \Omega_t^\tau \qquad \textit{s.t.} \qquad \sigma_t(z_t) \geq \frac{1}{\alpha} \max_{z \in \Omega_t^\tau} \sigma_t(z). \tag{48}$$

Crucially, the sampling oracle in (48) requires the generative model (e.g., via inference-time techniques) to sample approximate maximizers of $\sigma_t(\cdot)$ over the current generable set Ω_t^τ.

We ultimately want to show a suitable notion of expansion of the valid set by following the above procedure. As our valid set at any time is defined as an expansion of the valid set at the previous time, we cannot guarantee convergence to the

entire subset of $\mathcal{Z}$ which is valid. Instead, we must define some notion of *reachable valid set*. To this end, we consider the tightened one-step reachability operator

$$R_\varepsilon(S) \triangleq \{z \in \mathcal{Z} \mid \exists z' \in S : s(g(z')) - L_s L_g d(z, z') - \varepsilon \geq h\}.$$

The H-fold recursive application of this operator is denoted $R_\varepsilon^H(S)$, and its closure as $H \to \infty$ is denoted $\bar{R}_\varepsilon(S)$. Ultimately we show that the valid set discovered by the learner converges to $\bar{R}_0(S_0)$, leading to the following theorem.

Theorem E.5. *Fix some $\gamma > 0, \delta > 0$, $\varepsilon > 0$. Let H be a positive integer. Suppose that the sequence of calibration coefficients $\left\{\hat{\beta}_t(\delta)\right\}$ used to construct the sets S_t satisfies $\hat{\beta}_t(\delta) \geq \beta_t^{\Omega_* \cup \Gamma(\gamma)}(\delta)$. Let Assumptions E.1-E.2 hold at level γ and condition on the calibration event of Theorem E.1 with $A \leftarrow \Omega_* \cup \Gamma(\gamma)$. Let $\frac{\exp(\mathrm{logit}(\ell))}{\bar{Z}} \leq \tau \leq \frac{\exp(\mathrm{logit}(h)}{\bar{Z}}$. Consider sampling with the oracle defined by Assumption E.3 for $T^\star = TH$ steps, with T satisfying*

$$T \geq \frac{8\alpha^2 \hat{\beta}_{HT}(\delta)^2 \gamma_T^{\Omega_* \cup \Gamma(\gamma)}}{\varepsilon^2 \log(1 + (\lambda\kappa)^{-1})}.$$

Then it holds that $R_\varepsilon^H(S_0) \subseteq \Omega_{T^\star}^\tau$.

To show the above result, we first prove a few basic inequalities. Our first inequality bounds the uncertainty over the set of valid decisions after T additional algorithm steps.

Lemma E.6 (Uncertainty Reduction via Local Generative Sampling)**.** *Consider the setting of Lemma E.4. Fix $t \geq 0$ and $T \geq 1$. Under local generative sampling (48), it holds that the epistemic uncertainty at time $t + T$ satisfies*

$$\max_{z \in S_t} \sigma_{t+T}(z) \leq \alpha \sqrt{\frac{2}{\log(1 + (\lambda\kappa)^{-1})}} \sqrt{\frac{\gamma_T^{\Omega_* \cup \Gamma(\gamma)}}{T}}.$$

Proof. By monotonicity of the uncertainty score, it holds that

$$\begin{aligned}
\max_{z \in S_t} \sigma_{t+T}(z) &\leq \frac{1}{T} \sum_{n=0}^{T-1} \max_{z \in S_t} \sigma_{t+n}(z) && \text{(uncertainty is monotone)} \\
&\leq \frac{1}{T} \sum_{n=0}^{T-1} \max_{z \in S_{t+n}} \sigma_{t+n}(z) && (S_t \text{ are monotone}) \\
&\leq \frac{1}{T} \sum_{n=0}^{T-1} \max_{z \in \Omega_{t+n}^\tau} \sigma_{t+n}(z) && \text{(Lemma E.3)} \\
&\leq \frac{\alpha}{T} \sum_{n=0}^{T-1} \sigma_{t+n}(z_{t+n}) && \text{(local generative sampling (48))} \\
&\leq \frac{\alpha}{\sqrt{T}} \sqrt{\sum_{n=0}^{T-1} \sigma_{t+n}^2(z_{t+n})} && \text{(Cauchy–Schwarz).}
\end{aligned}$$

It holds that $x \leq \frac{\bar{x}}{\log(1+\bar{x})} \log(1+x)$ for $0 \leq x \leq \bar{x}$. Let $K_{t:t+T-1}$ be the Gram matrix of the queried points $z_t, \ldots, z_{t+T-1}$, let $K_{1:t-1}$ be the Gram matrix of the previously queried points $z_1, \ldots, z_{t-1}$, let $K_{t:t+T-1,\,1:t-1}$ be the corresponding cross-Gram matrix, and define

$$\Sigma_{t:t+T-1|1:t-1} \triangleq K_{t:t+T-1} - K_{t:t+T-1,\,1:t-1}(K_{1:t-1} + \lambda\kappa I_{t-1})^{-1} K_{1:t-1,\,t:t+T-1}.$$

Then the above inequality along with the fact that $k(z, z') \le 1$ for all $z, z' \in \Omega_* \cup \Gamma(\gamma)$ implies that

$$\begin{aligned}
\max_{z \in S_t} \sigma_{t+T}(z) &\le \frac{\alpha}{\sqrt{T}} \sqrt{\frac{1}{\log(1+(\lambda\kappa)^{-1})}} \sqrt{\sum_{n=0}^{T-1} \log\left(1 + \frac{\sigma_{t+n}^2(z_{t+n})}{\lambda\kappa}\right)} \\
&= \frac{\alpha}{\sqrt{T}} \sqrt{\frac{1}{\log(1+(\lambda\kappa)^{-1})}} \sqrt{\log\det(I_T + (\lambda\kappa)^{-1}\Sigma_{t:t+T-1|1:t-1})} \\
&\le \frac{\alpha}{\sqrt{T}} \sqrt{\frac{1}{\log(1+(\lambda\kappa)^{-1})}} \sqrt{\log\det(I_T + (\lambda\kappa)^{-1}K_{t:t+T-1})} \\
&\le \frac{\alpha}{\sqrt{T}} \sqrt{\frac{1}{\log(1+(\lambda\kappa)^{-1})}} \sqrt{\max_{z_1,\dots,z_T \in \Omega_* \cup \Gamma(\gamma)} \log\det(I_T + (\lambda\kappa)^{-1}K_T)} \\
&= \alpha\sqrt{2}\frac{\sqrt{\gamma_T^{\Omega_* \cup \Gamma(\gamma)}}}{\sqrt{T}} \sqrt{\frac{1}{\log(1+(\lambda\kappa)^{-1})}}.
\end{aligned}$$

Here the last inequality uses Lemma E.4, which implies that all queried points $z_t, \dots, z_{t+T-1}$ lie in $\Omega_* \cup \Gamma(\gamma)$. □

The next lemma shows that once the uncertainty over the current valid set is sufficiently small, additional local generative sampling expands the valid set according to the one-step reachability operator.

Lemma E.7 (One-step Reachable Expansion via Local Generative Sampling)**.** *Consider the setting of Lemma E.6. Let $\varepsilon > 0$, $\delta > 0$, $t \ge 0$, and $T \ge 1$. Suppose*

$$T \ge \frac{8\alpha^2 \hat{\beta}_{t+T}(\delta)^2 \gamma_T^{\Omega_* \cup \Gamma(\gamma)}}{\varepsilon^2 \log(1+(\lambda\kappa)^{-1})}.$$

Then, it holds that

$$R_\varepsilon(S_t) \subseteq S_{t+T}.$$

Proof. By Lemma E.6, the condition on T implies that

$$\max_{z \in S_t} \sigma_{t+T}(z) \le \frac{\varepsilon}{2\hat{\beta}_{t+T}}, \tag{49}$$

where we have adopted the shorthand $\hat{\beta}_t$ for $\hat{\beta}_t(\delta)$. Consider any point $z \in R_\varepsilon(S_t)$. By definition, there exists $z' \in S_t$ such that

$$h \le s(g(z')) - \varepsilon - L_s L_g d(z, z').$$

By the confidence event at round $t + T$,

$$s(g(z')) \le s(\mu_{t+T}(z')) + \hat{\beta}_{t+T}\sigma_{t+T}(z').$$

Hence

$$\begin{aligned}
h &\le s(g(z')) - \varepsilon - L_s L_g d(z, z') \\
&\le s(\mu_{t+T}(z')) + \hat{\beta}_{t+T}\sigma_{t+T}(z') - \varepsilon - L_s L_g d(z, z') \\
&\le s(\mu_{t+T}(z')) - \hat{\beta}_{t+T}\sigma_{t+T}(z') - L_s L_g d(z, z') \qquad (z' \in S_t, \text{ substitute (49)}).
\end{aligned}$$

Since the sets S_t are monotone, $z' \in S_t$ implies $z' \in S_{t+T-1}$. Therefore there exists $z' \in S_{t+T-1}$ such that

$$h \le s(\mu_{t+T}(z')) - \hat{\beta}_{t+T}\sigma_{t+T}(z') - L_s L_g d(z, z').$$

By the definition of S_{t+T} in (40), this implies $z \in S_{t+T}$. □

Applying the above result recursively leads to H-step expansion of S_0 at some timestep $T^\star$. In particular, suppose $T^\star = HT$ for some T satisfying

$$T \geq \frac{8\alpha^2 \hat{\beta}_{HT}(\delta)^2 \gamma_T^{\Omega_* \cup \Gamma(\gamma)}}{\varepsilon^2 \log(1 + (\lambda\kappa)^{-1})}.$$

This implies that the condition of the above lemma holds for each interval of length T, since γ_t and $\hat{\beta}_t$ are monotonically increasing. Consequently, it holds that

$$R_\varepsilon^H(S_0) \subseteq S_{T^\star}.$$

We can find $T^\star$ large enough to satisfy this condition as long as the complexity term $\hat{\beta}_t^2 \gamma_t^{\Omega_* \cup \Gamma(\tau)}$ grows sublinearly. Ultimately, by Lemma E.3, we have

$$S_{T^\star} \subseteq \Omega_{T^\star}^\tau,$$

which implies that the generable set $\Omega_{T^\star}^\tau$ obtained after running the algorithm is a superset of the valid reachable set, namely

$$R_\varepsilon^H(S_0) \subseteq S_{T^\star} \subseteq \Omega_{T^\star}^\tau. \tag{50}$$

While (50) states that the generable set of the extended diffusion model is a superset of the reachable valid set, it does not clarify whether enough model density has been placed within the valid reachable set $R_\varepsilon^H(S_0) \subseteq S_{T^\star}$ rather than within $\Omega_{T^\star}^\tau \setminus R_\varepsilon^H(S_0)$, which might contain invalid points. To shed light on this question, we next derive a pointwise lower bound on the final density over the generable set.

E.4. Deriving a Lower Bound on Validity

We now show that samples from the generative model trained after $T^\star$ are valid with high probability.

Corollary E.8. *Under the setting of Theorem E.5 and by the construction of the EBM model, it holds that*

$$\mathbb{P}_{z \sim p_{T^\star}}[z \in \Omega_*] \geq \mathsf{vol}(R_\varepsilon^H(S_0)) \frac{\exp(\mathrm{logit}(h))}{\bar{Z}}.$$

Proof. It holds by definition of $p_{T^\star}$ that

$$\begin{aligned}
\mathbb{P}_{z \sim p_{T^\star}}[z \in \Omega_*] &= \int_{z \in \mathcal{Z}} \mathbf{1}(\Omega_*)(z) p_{T^\star}(z) dz \\
&= \int_{z \in \mathcal{Z}} \mathbf{1}(\Omega_*)(z) \frac{\exp(\mu'_{T^\star}(z))}{\int_{\mathcal{Z}} \exp(\mu'_{T^\star}(z)) dz} dz \\
&\geq \int_{z \in R_\varepsilon^H(S_0)} \mathbf{1}(\Omega_*)(z) \frac{\exp(\mu'_{T^\star}(z))}{\bar{Z}} dz \\
&= \int_{z \in R_\varepsilon^H(S_0)} \frac{\exp(\mu'_{T^\star}(z))}{\bar{Z}} dz \\
&\geq \int_{z \in R_\varepsilon^H(S_0)} \frac{\exp(\mathrm{logit}(h))}{\bar{Z}} dz,
\end{aligned}$$

where the last equality follows from the fact that $R_\varepsilon^H(S_0) \subseteq \Omega_*$, and the last inequality follows from the fact that $R_\varepsilon^H(S_0) \subseteq S_{T^\star}$ and $\mu'_{T^\star}(z) \geq \mathrm{logit}(h)$ for $z \in S_{T^\star}$. □

E.5. Proof of design-space coverage corollary

Assumption E.4 (Fixed representation with nondegenerate Jacobian)**.** *The representation map $\phi : \mathcal{X} \to \mathcal{Z}$ is fixed throughout ADE, is a C^1-diffeomorphism onto $\mathcal{Z} = \phi(\mathcal{X})$, and satisfies $|\det J\phi(x)| \geq j_{\min} > 0$ for all $x \in \mathcal{X}$.*

Corollary 4.2 (Design-space coverage of the induced reachable valid set)**.** *Assume the conditions of Theorem 4.1 and Assumption E.4, with $j_{\min} := \inf_{x \in \mathcal{X}} |\det J_\phi(x)| > 0$. Let $S_0^X := \phi^{-1}(S_0)$ and $\tau_X := j_{\min}\tau$. Then, after the same number $T^\star$ of verified samples as in Theorem 4.1, with probability at least $1 - \delta$,*

$$\left(R_\epsilon^{X,\phi}\right)^H (S_0^X) \subseteq \Omega_{T^\star}^{X,\tau_X}. \tag{22}$$

Proof. Condition on the event of Theorem 4.1, which holds with probability at least $1-\delta$. On this event,

$$R_\epsilon^H(S_0) \subseteq \Omega_{T^\star}^\tau = \{z \in \mathcal{Z} : \ p'_{T^\star}(z) \geq \tau\}. \tag{51}$$

We first show that for every $A \subseteq \mathcal{X}$,

$$\phi\big(R_\epsilon^{X,\phi}(A)\big) = R_\epsilon\big(\phi(A)\big). \tag{52}$$

Indeed, if $x \in R_\epsilon^{X,\phi}(A)$, then for some $x' \in A$,

$$s\big(g(\phi(x'))\big) - L_s L_g \, d\big(\phi(x), \phi(x')\big) - \epsilon \geq h.$$

Writing $z = \phi(x)$ and $z' = \phi(x')$, we get $z' \in \phi(A)$ and

$$s(g(z')) - L_s L_g \, d(z, z') - \epsilon \geq h,$$

hence $z \in R_\epsilon(\phi(A))$. This proves $\phi(R_\epsilon^{X,\phi}(A)) \subseteq R_\epsilon(\phi(A))$.

Conversely, if $z \in R_\epsilon(\phi(A))$, then for some $z' \in \phi(A)$,

$$s(g(z')) - L_s L_g \, d(z, z') - \epsilon \geq h.$$

Since ϕ is bijective, there exist unique $x = \phi^{-1}(z)$ and $x' = \phi^{-1}(z')$ with $x' \in A$. Substituting $z = \phi(x)$ and $z' = \phi(x')$ gives

$$s\big(g(\phi(x'))\big) - L_s L_g \, d\big(\phi(x), \phi(x')\big) - \epsilon \geq h,$$

so $x \in R_\epsilon^{X,\phi}(A)$ and thus $z = \phi(x) \in \phi(R_\epsilon^{X,\phi}(A))$. Therefore (52) holds.

Iterating (52) yields

$$\phi\Big(\big(R_\epsilon^{X,\phi}\big)^H(S_0^X)\Big) = R_\epsilon^H\big(\phi(S_0^X)\big) = R_\epsilon^H(S_0), \tag{53}$$

where we used $S_0^X = \phi^{-1}(S_0)$.

Now let $x \in \big(R_\epsilon^{X,\phi}\big)^H(S_0^X)$. By (53), its image $z := \phi(x)$ belongs to $R_\epsilon^H(S_0)$, hence by (51),

$$p'_{T^\star}(z) \geq \tau. \tag{54}$$

Since ϕ is a C^1-diffeomorphism, the change-of-variables formula implies that the design-space and representation-space densities satisfy

$$p_1^{\pi_{T^\star}}(x) = p'_{T^\star}(\phi(x)) \, |\det J\phi(x)|.$$

Combining this with (54) and the Jacobian lower bound from Assumption E.4 gives

$$p_1^{\pi_{T^\star}}(x) \geq \tau \, j_{\min} = \tau_X.$$

Therefore $x \in \Omega_{T^\star}^{X,\tau_X}$ by (20). Since x was arbitrary, (22) follows. □

E.6. Full design-space coverage under global reachability

Corollary E.9 (Full coverage under finite-chain global reachability)**.** *Assume the conditions of Theorem E.5 and Assumption E.4. Let*

$$S_0^X := \phi^{-1}(S_0), \qquad \tau_X := j_{\min}\tau,$$

as in Corollary 4.2. Suppose that for every design $x \in \Omega_*$ there exists a finite chain

$$x_0, \ldots, x_m \in \Omega_*, \qquad x_0 \in S_0^X, \quad x_m = x, \quad m \leq H,$$

such that for every $i = 1, \ldots, m$,

$$L_s L_g \, d\big(\phi(x_i), \phi(x_{i-1})\big) + \varepsilon \ \leq \ s\big(g(\phi(x_{i-1}))\big) - h.$$

Then

$$\Omega_* \subseteq \Omega_{T^\star}^{X,\tau_X}$$

with probability at least $1-\delta$, after the same number $T^\star = HT$ of verified samples as in Theorem E.5.

Proof. Condition on the calibration event used in Theorem E.5, which holds with probability at least $1-\delta$ by Theorem E.1 under the conditions of Theorem E.5. Fix $x \in \Omega_*$, and let $x_0, \dots, x_m$ be the chain given by the assumption. We first prove by induction that

$$x_i \in \left(R_\varepsilon^{X,\phi}\right)^i (S_0^X) \qquad \text{for all } i \le m,$$

with the convention $\left(R_\varepsilon^{X,\phi}\right)^0 (S_0^X) = S_0^X$. The base case $i = 0$ is immediate since $x_0 \in S_0^X$. Assume $x_{i-1} \in \left(R_\varepsilon^{X,\phi}\right)^{i-1} (S_0^X)$. By the chain condition,

$$s\big(g(\phi(x_{i-1}))\big) - L_s L_g \, d\big(\phi(x_i), \phi(x_{i-1})\big) - \varepsilon \ge h.$$

Therefore, by the definition of $R_\varepsilon^{X,\phi}$ in (21),

$$x_i \in R_\varepsilon^{X,\phi}\left(\left(R_\varepsilon^{X,\phi}\right)^{i-1}(S_0^X)\right) = \left(R_\varepsilon^{X,\phi}\right)^i (S_0^X).$$

Thus

$$x = x_m \in \left(R_\varepsilon^{X,\phi}\right)^m (S_0^X).$$

Let $z = \phi(x)$. Iterating the intertwining identity (52) m times gives

$$\phi\left(\left(R_\varepsilon^{X,\phi}\right)^m (S_0^X)\right) = R_\varepsilon^m\big(\phi(S_0^X)\big) = R_\varepsilon^m(S_0),$$

where the last equality uses $S_0^X = \phi^{-1}(S_0)$. Hence

$$z \in R_\varepsilon^m(S_0).$$

We now show that every intermediate exact iterate is covered by the certified set at the corresponding block time. Specifically, for every $r \le H$,

$$R_\varepsilon^r(S_0) \subseteq S_{rT}.$$

The claim holds for $r = 0$. If it holds for some $r < H$, then by monotonicity of the operator R_ε and by Lemma E.7, applied on the block starting at time rT,

$$R_\varepsilon^{r+1}(S_0) = R_\varepsilon\big(R_\varepsilon^r(S_0)\big) \subseteq R_\varepsilon(S_{rT}) \subseteq S_{(r+1)T}.$$

Here the sample-complexity condition in Theorem E.5 ensures that Lemma E.7 applies on every block of length T. Since $m \le H$, the preceding inclusion and the monotonicity of the certified sets S_t, which follows from their recursive definition in (40), imply

$$z \in R_\varepsilon^m(S_0) \subseteq S_{mT} \subseteq S_{HT} = S_{T^\star}.$$

By Lemma E.3,

$$S_{T^\star} \subseteq \Omega_{T^\star}^\tau.$$

Therefore $p'_{T^\star}(\phi(x)) \ge \tau$. Finally, by the change-of-variables formula and Assumption E.4,

$$p_1^{\pi_{T^\star}}(x) = p'_{T^\star}(\phi(x)) \, |\det J\phi(x)| \ge \tau j_{\min} = \tau_X.$$

Hence $x \in \Omega_{T^\star}^{X,\tau_X}$. Since $x \in \Omega_*$ was arbitrary,

$$\Omega_* \subseteq \Omega_{T^\star}^{X,\tau_X}.$$

□

F. ActFlow for Discrete Diffusion

To adapt ACTFLOW for *discrete diffusion models*, we leverage the algorithm introduced in prior work for adapting a pre-trained discrete diffusion model to an intractable reward-tilted distribution (Tang et al., 2025). Given a reward function defined over clean sequences, this method provably tilts a pre-trained discrete diffusion model to the reward-tilted distribution using off-policy reinforcement learning. The full algorithm is given in Alg 2.

F.1. Discrete Diffusion as Continuous-Time Markov Chains

In contrast to the continuous state space, where generative models are defined by stochastic differential equations (SDEs) or ordinary differential equations (ODEs), generative models in the discrete state space $\mathcal{V} := \{1, \dots V\}$ are defined by **continuous-time Markov chains** (CTMCs). A CTMC is a stochastic process $X_{0:1} := (X_s)_{s\in[0,1]}$ whose probability law is defined by a *generator* $(\boldsymbol{Q}_s)_{s\in[0,1]} \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ defined as:

$$\boldsymbol{Q}_s(x, y) = \lim_{\Delta s\to 0} \frac{1}{\Delta s}(\Pr(X_{s+\Delta s} = y|X_s = x) - \mathbf{1}_{x=y}) \tag{55}$$

which defines the probability of transitioning from state $x \in \mathcal{V}$ to state $y \in \mathcal{V}$ at time s.

Masked discrete diffusion models (MDMs) (Shi et al., 2024; Sahoo et al., 2024; Ou et al., 2024; Zheng et al., 2024) are a class of discrete diffusion models that aim to generate sequences $x_1 \sim p_{\text{data}} \in \mathcal{V}^L$ of length L from a sequence of absorbing *mask tokens* $\boldsymbol{M}$. The generative process is defined as a CTMC that evolves from a prior distribution p_0 defined as the Dirac delta of fully masked sequences to the data distribution $p_1 \equiv p_{\text{data}}$. Since each position along the sequence $\ell \in \{1, \dots, L\}$ transitions from a mask token $X^\ell = \boldsymbol{M}$ to a clean token $X^\ell = x^\ell$, the generator can be parameterized as:

$$\boldsymbol{Q}_s(x, x^{\ell\leftarrow v}) = \gamma(s)p^{\theta_t}(\cdot|x)_{\ell,v} \tag{56}$$

where $x^{\ell\leftarrow v}$ denotes the sequence where the ℓth token is replaced with state $v \in \mathcal{V}$ and $\gamma(t)$ is the forward noising schedule. To train $p^{\theta_t}(\cdot|x)$ to reconstruct sequences from the data distribution, we can optimize a **denoising cross-entropy (DCE)** loss (Shi et al., 2024; Sahoo et al., 2024; Ou et al., 2024) defined as:

$$\mathcal{L}_{\text{DCE}}(\theta; x_1) := \mathbb{E}_{s\sim\mathcal{U}(0,1)}\left[\frac{1}{s}\mathbb{E}_{p_s(\tilde{x}_s|x_1)}\sum_{\ell:\tilde{x}_s^\ell=\boldsymbol{M}} -\log p^{\theta_t}(\cdot|\tilde{x}_s^\ell)_\ell\right], \quad x_1 \sim p_{\text{data}} \tag{57}$$

where $p_s(\tilde{x}_s|x_1)$ is the distribution of partially masked sequences obtained by masking each position with probability s given a clean sequence $x_1 \sim p_{\text{data}}$.

F.2. Entropy-Regularized Uncertainty Optimization for Discrete Diffusion

Given a function $\sigma(\cdot) : \mathcal{V}^L \to \mathbb{R}$ that returns the epistemic uncertainty of a sequence $x_1 \in \mathcal{V}^L$ and a pretrained discrete diffusion model that generates the path measure $\mathbb{P}^{\theta_0}$, we define the **uncertainty-tilted path measure** as:

$$\mathbb{P}^\sigma(X_{0:1}) := \frac{1}{Z}\mathbb{P}^{\theta_0}(X_{0:1})\exp\left(\frac{\sigma(X_1)}{\beta}\right), \quad p_1^\sigma(X_1) \propto p_{\text{data}}(X_1)\exp\left(\frac{\sigma(X_1)}{\beta}\right) \tag{58}$$

This uncertainty-tilted path measure coincides with the solution to the **entropy-regularized reward optimization** problem with reward defined as the uncertainty $r(\cdot) := \sigma(\cdot)$ given by:

$$\arg\max_{\theta_{t+1}} \mathbb{E}_{X_{0:1}\sim\mathbb{P}^{\theta_t}}[\sigma(X_1)] - \beta D_{\text{KL}}(\mathbb{P}^{\theta_t}\|\mathbb{P}^{\theta_0}) \tag{59}$$

where $\mathbb{P}^{\theta_t}$ is the CTMC path measure generated from the adapted model with parameters θ and $\mathbb{P}^{\theta_0}$ is the frozen pre-trained model. β is the weight of the KL regularization with the pre-trained model.

F.3. Uncertainty-Aware Fine-Tuning of Discrete Diffusion

To adapt a pre-trained discrete diffusion model to align with the *uncertainty-tilted distribution* defined in (58), we leverage the off-policy reinforcement learning algorithm introduced in prior work (Tang et al., 2025), which is minimized when

Algorithm 2 ACTFLOW for Discrete Diffusion

Input: pre-trained model $p^{\theta_0}(\cdot|X_s)$, uncertainty function $\sigma : \mathcal{V}^L \to \mathbb{R}$, number of iterations T, number of WDCE repeats R
for $t = 0, 1, \ldots, T-1$ **do**
  $\{x_1^i, w^\sigma\}_{i=1}^B \leftarrow \texttt{Generate}(p^{\theta_0}, p^{\theta_t})$
  $\mathcal{B} \leftarrow \{x_1^i, w^\sigma\}_{i=1}^B$ {replay buffer}
  for step in $1, \ldots, N_{\text{step}}$ **do**
    $\{\tilde{x}_s^i, w^\sigma\}_{i=1}^{B\times R} \leftarrow \texttt{ResampleWithMask}(\mathcal{B}; R)$
    Compute $\mathcal{L}_{\text{WDCE}}$ from (60) with $\{\tilde{x}_s^i, w^\sigma\}_{i=1}^{B\times R}$
    Update θ_{t+1} with $\nabla_\theta \mathcal{L}_{\text{WDCE}}$
  end for
end for

$\mathbb{P}^{u_\theta} = \mathbb{P}^\sigma$. The training objective is defined as the **weighted denoising cross-entropy** loss $\mathcal{L}_{\text{WDCE}}$ given by:

$$\mathcal{L}_{\text{WDCE}} := \mathbb{E}_{X_{0:1}\sim\mathbb{P}^\sigma}\left[\mathcal{L}_{\text{DCE}}(\theta; x_1)\right] = \mathbb{E}_{X_{0:1}\sim\mathbb{P}^{\bar{u}}}\left[\underbrace{\frac{\mathrm{d}\mathbb{P}^\sigma}{\mathrm{d}\mathbb{P}^{\bar{\theta}_t}}}_{\exp(w^\sigma(X_1))}\mathcal{L}_{\text{DCE}}(\theta; x_1)\right] \tag{60}$$

where $\mathbb{P}^{\bar{\theta}_t} = \texttt{stopgrad}(\mathbb{P}^{\theta_t})$ is the CTMC path measure generated from the non-gradient-tracking model. We define the importance weight $w^\sigma(X_1) := \log \frac{\mathrm{d}\mathbb{P}^\sigma}{\mathrm{d}\mathbb{P}^{\bar{\theta}_t}}$ as the Radon-Nikodym derivative between the uncertainty-tilted path measure $\mathbb{P}^\sigma$ and the non-gradient-tracking adapted model $\mathbb{P}^{\bar{\theta}_t}$:

$$\log \frac{\mathrm{d}\mathbb{P}^\sigma}{\mathrm{d}\mathbb{P}^{\bar{\theta}_t}}(X_{0:1}) = \underbrace{\frac{\sigma(X_1)}{\beta} + \sum_{s:X_{s+\Delta s}\neq X_s}\sum_{\ell:X^\ell_{s+\Delta s}\neq X^\ell_s}\log\frac{p^{\theta_0}(X^\ell_{s+\Delta s}|X_s)}{p^{\bar{u}}(X^\ell_{s+\Delta s}|X_s)}}_{w^\sigma(X_{0:1})} - \log Z \tag{61}$$

which reweights trajectories $X_{0:1}$ from the current frozen model by their likelihood under the uncertainty-tilted path measure. To optimize this objective, we iterate through the following steps:

(i) Sample B trajectories $X_{0:1}^i$ from the adapted model $\mathbb{P}^{\bar{\theta}_t}$ without gradient tracking while tracking the log-likelihoods of each step $\log p^{\theta_0}(X_{s+\Delta s}|X_s) - \log p^{\bar{\theta}_t}(X_{s+\Delta s}|X_s)$.

(ii) Compute the importance weights $w^\sigma(X_{0:1}^i)$ using the log-likelihoods and the uncertainties evaluated on the clean sequences $\sigma(X_1^i)$.

(iii) Store the clean sequences and their weights in a replay buffer $\mathcal{B} \leftarrow \{x_1^i, w^\sigma(x_1^i)\}_{i=1}^B$.

(iv) Resample R partially masked versions of each clean sequence in the buffer x_1^i at different time steps $s \in [0, 1]$ to obtain $\{\tilde{x}_s^i, w^\sigma(x_1^i)\}_{i=1}^{B\times R}$.

(v) Compute $\mathcal{L}_{\text{WDCE}}(\theta)$ from Eq (60) using $\{\tilde{x}_s^i, w^\sigma(x_1^i)\}_{i=1}^{B\times R}$ and update θ with $\nabla_\theta \mathcal{L}_{\text{WDCE}}$ for N_{iter} iterations.

G. Experimental Details

G.1. Domain-agnostic evaluation metrics for real-world OOD generative modeling.

Standard evaluation criteria for deep generative models, such as Frechet Inception Distance (FID) (Heusel et al., 2017), aim to assess whether the generative model well-approximates the data distribution p_{data}. This is misaligned with OOD generative modeling: successful expansion should increase valid coverage beyond the initial generable region, and may therefore *increase* distributional distance from the pre-trained model. In low-dimensional illustrative experiments, valid coverage can be estimated directly by discretizing the design space, together with validity. In molecular and protein spaces, however, such direct coverage computation is infeasible. We therefore employ several domain-agnostic evaluation metrics for OOD generative modeling, namely: (i) number of fixed-threshold clusters covered by valid samples, to measure model coverage; (ii) Vendi diversity (Friedman & Dieng, 2022), to measure distributional diversity; (iii) FID to quantify divergence from the pre-trained model distribution, which we wish to increase; and validity percentage, to assess whether model expansion is happening over valid regions. In particular, the number of clusters corresponds to the number of hyper-spheres with a fixed data-specific radius, forming a packing over a finite draw of samples of fixed size across methods. This allows to approximately describe the volume covered by the union of hyper-spheres centered at the generated data points, and is an efficient-to-compute proxy metric to assess coverage of the model generable set.

G.2. Illustrative 2D Experiments

Overview. We evaluate ACTFLOW in a two-dimensional illustrative design space. The base model is a continuous flow model over $\mathbb{R}^2$. The initial model is deliberately misspecified by centering the pretraining data at $(-1.1, 0)$, using 512 Gaussian samples with standard deviation 0.1. The valid region is a 3×3 checkerboard over $[-3.5, 3.5]^2$: a point is valid if it lies in-bounds and falls in a checkerboard cell with even parity. The base model is pretrained for 2500 steps with Adam, learning rate 10^{-3}, and batch size 256. Results are acquired over 20 seeds and 95% CIs are shown.

Algorithm configuration. We run ACTFLOW for 500 iterations and 20 random seeds. At each iteration, ACTFLOW self-generates 64 samples. Fine-tuning uses batch size 256 and 250 gradient steps per iteration. Evaluation is performed every 50 iterations using 3000 samples for evaluation curves. We use $\beta = 1/13$ and GP with RBF kernel with lengthscale 0.08, and negative-gradient scale $\alpha_t = 0.005$.

Uncertainty estimation. We use a Gaussian process with RBF kernel lengthscale set to 0.08 and employ flow representation timestep $s = 0.9$

Validity estimation (i.e., verifier). The verifier is the deterministic checkerboard validity function. The domain $[-3.5, 3.5]^2$ is partitioned into 3×3 equal cells. A sample is valid if it lies inside the domain and its cell index (i, j) satisfies $(i + j) \bmod 2 = 0$.

Coverage metric. Coverage is measured as generable coverage over the valid region. We draw $3{,}000$ samples from the current model and estimate its generable set on a 100×100 histogram over $[-3.5, 3.5]^2$. A bin is considered generable if its estimated density is at least $\tau = 0.01$.

Ablations: no negative gradient in flow matching loss. We report in Fig. 7 visual results for the illustrative experiments, run with same parameters as in 2, except for α_t, which is now set to 0.0. As one can notice from Fig. 7, the change in this parameters leads to seemingly minimal changes in the results according to the metrics assessed.

Baselines. Both baselines use the same number of iterations, samples per iteration, fine-tuning steps, fine-tuning batch size, evaluation schedule, initial invalid model setting, and seeds as ACTFLOW.

Hardware and Compute. Each 2D experiment job requests one RTX 2080 GPU, 16GB memory per CPU, and a four-hour wall-clock limit.

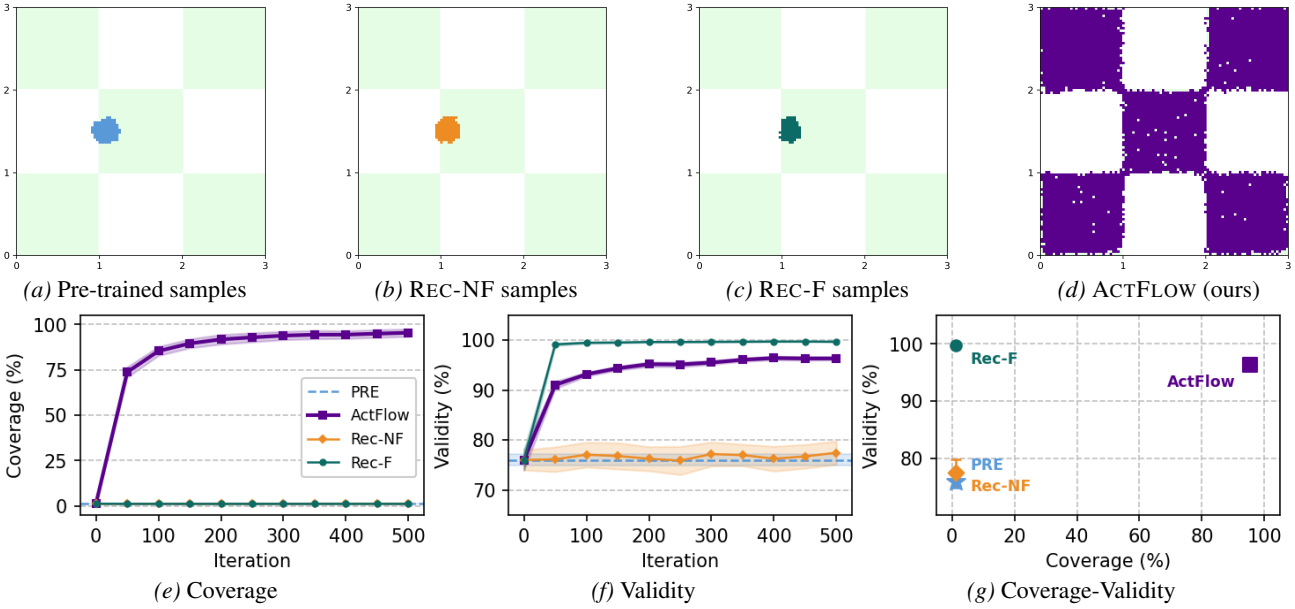


Figure 7. Results for illustrative experiments with same parameters as in 2, except for $\alpha_t = 0$.

G.3. Molecular Design: QM9 Experiments

Overview. We apply our method to small-molecule generation using FlowMol Gaussian (Dunn & Koes, 2024) pre-trained on QM9 (Ramakrishnan et al., 2014). Results are acquired over 5 seeds and 95% CIs are shown.

Algorithm configuration. We run ACTFLOW for 1,066 iterations, with 66 initial warm-up iterations during which the model is not fine-tuned. Each iteration consists of 64 samples, followed by 500 fine-tuning gradient steps. Fine-tuning uses AdamW with learning rate 10^{-4}, and batch size 64. Fine-tuning is deferred until 4096 valid samples have been collected, after which the accumulated buffer is used jointly with new guided samples at each iteration. We employ $\beta = 1/10 = 0.1$, and $s = 0.9$.

Uncertainty estimation. We use a deep bootstrapped ensemble of 5 MLPs, each with two hidden layers of 100 units, ReLU activations, and 10% dropout. Each ensemble member is trained independently on a 90% bootstrap subsample of the accumulated feature label pairs, using Adam with learning rate 10^{-3}, for up to 1000 steps. The ensemble standard deviation across members is used as the uncertainty signal.

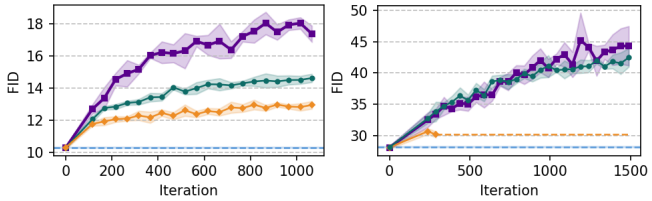
Validity estimation (i.e., verifier). A generated molecule is deemed valid if its RDKit-sanitised representation passes valence and bond-order checks and consists of a single connected fragment. Sanitisation is preceded by an MMFF geometry relaxation (Halgren, 1996).

Coverage metric. Coverage is measured as the number of distinct molecular clusters obtained, computed via greedy sphere exclusion on Morgan fingerprints (radius 2, 2048 bits) using Tanimoto similarity with threshold $\tau = 0.85$: a candidate is added as a new cluster centre if its Tanimoto similarity to all previously selected centres is below τ. This is applied independently to the 500 valid molecules in each evaluation batch.

Diversity metric. Vendi score (Friedman & Dieng, 2022) is computed on the same 500 valid molecules per evaluation. The kernel matrix K is the pairwise Tanimoto similarity over 2048-bit Morgan fingerprints (radius 2).

Baselines. Both baselines use the same number of iterations, samples per iteration, fine-tuning steps, fine-tuning batch size, evaluation schedule, initial invalid model setting, and seeds as ACTFLOW.

FID We report in Fig. 8 FID results over iterates for QM9.



(a) QM9 FID *(b)* GEOM-Drugs FID

Figure 8. FID results for QM9 and GEOM-Drugs

Hardware and Compute. Each run used a single NVIDIA RTX 4090 GPU with 4 CPU cores and 32 GB of system memory, allocated for up to 120 h.

G.4. Molecular Design: GEOM-Drugs Experiments

Overview. We apply our method to small-molecule generation using FlowMol Gaussian (Dunn & Koes, 2024) pre-trained on GEOM-Drugs (Axelrod & Gomez-Bombarelli, 2022), a dataset of ∼300 k drug-like organic molecules with energy-annotated conformers. Results are acquired over 3 seeds and 95% CIs are shown.

Algorithm configuration. We run ActFlow for 1,500 iterations, with 190 initial warm-up iterations during which the model is not fine-tuned. Each iteration consists of 64 samples, followed by 2000 fine-tuning gradient steps. Fine-tuning uses AdamW with learning rate 10^{-4}, and effective batch size 64. Fine-tuning is deferred until 4096 valid samples have been collected. We employ $\beta = 1/7 \approx 0.14$, and $s = 0.8$.

Uncertainty estimation. We use a deep bootstrapped ensemble of 5 MLPs, each with two hidden layers of 100 units, ReLU activations, and 10% dropout. Each ensemble member is trained independently on a 90% bootstrap subsample of the accumulated feature–label pairs, using Adam with learning rate 10^{-3}, for up to 1000 steps. The ensemble standard deviation across members is used as the uncertainty signal.

Validity estimation (i.e., verifier). A generated molecule is deemed valid if its RDKit-sanitised representation passes valence and bond-order checks and consists of a single connected fragment. Sanitisation is preceded by an MMFF geometry relaxation (Halgren, 1996).

Coverage metric. Coverage is measured as the number of distinct molecular clusters obtained, computed via greedy sphere exclusion on Morgan fingerprints (radius 2, 2048 bits) using Tanimoto similarity with threshold $\tau = 0.85$: a candidate is added as a new cluster centre if its Tanimoto similarity to all previously selected centres is below τ. This is applied independently to the 500 valid molecules in each evaluation batch.

Diversity metric. Vendi score (Friedman & Dieng, 2022) is computed on the same 500 valid molecules per evaluation. The kernel matrix K is the pairwise Tanimoto similarity over 2048-bit Morgan fingerprints (radius 2).

Baselines. Both baselines use the same number of iterations, samples per iteration, fine-tuning steps, fine-tuning batch size, evaluation schedule, initial invalid model setting, and seeds as ActFlow.

FID We report in Fig. 8 FID results over iterates for GEOM-Drugs.

Hardware and Compute. Each run used a single NVIDIA RTX 4090 GPU with 4 CPU cores and 128 GB of system memory, allocated for up to 120 h.

G.5. Therapeutic Peptide Design Experiments

Overview. We apply ActFlow for therapeutic peptide design by adapting the pre-trained discrete diffusion peptide SMILES generator from PepTune (Tang et al., 2025) trained on 11 million peptide SMILES. The Simplified Molecular-Input Line-Entry System (SMILES) (Weininger, 1988) representation enables the generation of non-natural amino acids containing diverse chemical backbone and side-chain modifications, and cyclic modifications, significantly expanding the design space

of therapeutic peptides over the standard 20 natural amino acids. The tokenization uses the SMILES Pair Encoding (SPE) tokenization scheme (Li & Fourches, 2021) from PeptideCLM (Feller & Wilke, 2025) with vocabulary size 586, including 5 special tokens. Results are acquired over 5 seeds and 95% CIs are shown.

Algorithm configuration. We run ACTFLOW for 5 rounds following Alg 2, with an initial pool of 1000 peptide sequences generated from the pretrained model and 100 iterations per round. At each round, we draw 100 sequences from the pool, score them with the verifier and Gaussian process uncertainty model, update the policy with the WDCE loss defined in Apx F.3 (16 replicates per sequence), and refresh the training pool by sampling from the updated policy. We train with Adam at learning rate 1×10^{-4} and batch size 100. The reward is scaled by a parameter $\beta = 0.005$.

Uncertainty estimation. Sequence-level uncertainty is obtained from a Gaussian process (GP) fit on top of the diffusion model's RoFormer encoder. For each peptide, we extract attention-pooled, L_2-normalised hidden states from the final transformer layer (768-dimensions) and treat the GP posterior variance at that point as the uncertainty signal. The GP uses an RBF kernel length-scale is initialised from the mean pairwise embedding distance of 50 samples from the pretrained model. The GP's posterior is refit at each iteration on all sampled sequences.

Validity estimation (i.e., verifier). Generated SMILES are first parsed with RDKit. Those that yield a valid Mol object are then decoded with ther `SMILES2PEPTIDE` verifier (Tang et al., 2025) which decodes the SMILES into a sequence of natural and non-natural amino acids split on their peptide bonds.

Coverage metric. Coverage is measured as the fraction of generated peptides that occupy distinct chemical neighbourhoods, using sphere-exclusion clustering on Morgan/ECFP4 fingerprints (radius 2, 2048 bits). We report the number of clusters at a Tanimoto-similarity threshold of 0.001.

Diversity metric. We compute the Vendi score (Friedman & Dieng, 2022) computed using the RBF kernel on the pretrained RoFormer's L_2-normalised hidden states.

Baselines. We compare against three baselines: (1) the pretrained model (Tang et al., 2025), (2) REC-NF where the policy is updated on its own samples with no uncertainty tilting (uniform weights on WDCE loss), (3) REC-F where the policy is updated on its own samples filtered to retain only valid peptides classified by the verifier. We hold hyperparameters fixed across all baselines.

Hardware and Compute. Each mode and seed run is trained on a single NVIDIA B200 GPU with 8 CPU cores and 80 GB of system RAM. A full 100-iteration run fits within a 48-hour wallclock budget per GPU.

G.6. Protein Sequence Design Experiments

Overview. We apply our method to protein sequence design using a continuous ESM diffusion model from SGPO (Yang et al., 2025). The base model is a continuous-space denoising network operating over ESM token-probability vectors of dimension equal to the vocabulary size (31 tokens, including the 20 standard amino acids and special tokens), pre-trained on the CreiLOV fluorescence dataset (Chen et al., 2023). CreiLOV is a 119-residue fluorescent protein; the dataset contains experimentally measured fluorescence fitness values for sequence variants. The diffusion process uses a cosine noise schedule, where $\alpha_t = \cos\left(\frac{(1-t)\pi}{2}\right)$ and $\beta_t = \sqrt{1-\alpha_t^2}$, so that $t = 1$ corresponds to data and $t = 0$ to pure noise.

Algorithm configuration. We run ACTFLOW for 512 iterations, each consisting of 64 samples, followed by 1000 fine-tuning gradient steps on the accumulated valid samples. Fine-tuning uses AdamW with learning rate 10^{-4}, batch size 64, and no weight decay. Fine-tuning is deferred until 4096 valid samples have been collected (warm-up period), after which the accumulated buffer is used jointly with new guided samples at each iteration. We use $\beta = 1/50$. Features for the uncertainty estimator are extracted from the encoder of the ESM network at flow representation timestep $s = 0.8$, mean-pooled over the sequence dimension.

Uncertainty estimation. We use a deep ensemble of 5 MLPs, each with two hidden layers of 100 units, ReLU activations, and 10% dropout. Each ensemble member is trained independently on a 90% bootstrap subsample of the accumulated

feature-label pairs, using Adam with learning rate 10^{-3}, for up to 1000 steps. The ensemble standard deviation across members is used as the uncertainty signal.

Validity estimation (i.e., verifier). A generated sequence is deemed valid if its mean predicted local distance difference test (pLDDT), computed by ESMFold (Lin et al., 2022), exceeds a threshold of 65. ESMFold is run in batches of 32 sequences.

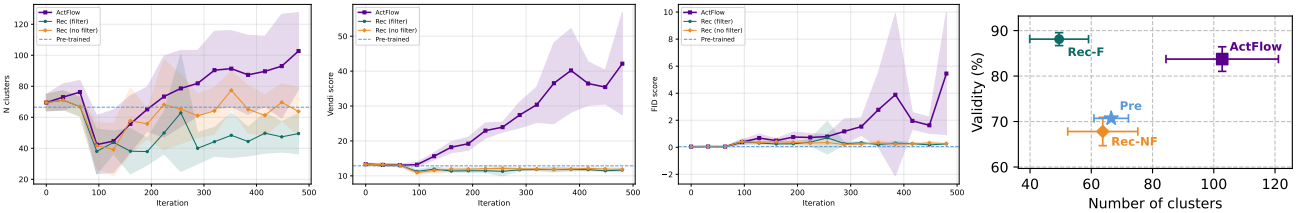
Coverage metric. Coverage is measured as the number of distinct sequence clusters accumulated across all 512 iterations, computed via greedy sphere exclusion on sequence identity with threshold $\tau = 0.35$: a candidate is added as a new cluster center if its sequence identity to all previously selected center is below τ.

Diversity metric. Vendi score (Friedman & Dieng, 2022) is computed on token-level ESM embeddings. For each generated sequence, we compute its embedding as the probability vector over the vocabulary projected through the ESM token embedding table (L2-normalised and scaled by $\sqrt{d_{\text{model}}}$), then mean-pooled over the sequence length. Vendi score is computed using an RBF kernel with lengthscale $\ell = 2.0$ applied to these mean-pooled embeddings.

Ablation: feature timestep. We ablated the fine-tuning flow representation time-steps $t \in \{0.5, 0.8, 0.9, 0.95\}$ in an alternative configuration with 1024 valid samples are initially queried. Among the feature timestep variants, $t = 0.95$ performed worst in terms of coverage (12 clusters at $\lambda = 50$, Vendi $= 18.6$), showing that extracting features close to the data level might be significantly sub-optimal.

Baselines. REC-F runs the same fine-tuning loop without uncertainty-guided sampling, training only on valid samples (pLDDT > 65). REC-NF additionally disables the validity filter during fine-tuning, training on all generated samples regardless of pLDDT. Both baselines use identical hyperparameters (fine-tuning steps, learning rate, batch size, warm-up threshold) to ACTFLOW, differing only in whether uncertainty guidance and validity filtering are applied.

Hardware and Compute. Each run used a single RTX 4090 GPU for 24 hours and 256 GB of memory.



(a) CreiLOV Coverage *(b)* CreiLOV Diversity *(c)* CreiLOV FID *(d)* Coverage-Validity

Figure 9. Results of protein sequence design experiments over iterations (Figs 9a - 9c), and diversity-validity tradeoff at final iteration (Fig 9d). ACTFLOW significantly outperforms REC-NF and REC-F in all diversity metrics (FID, Vendi, number of clusters), while maintaining a competitive validity.