LEARNING INFERENCE CLOCKS FOR FINITE-STEP GENERATIVE ODES

Anonymous authors
Paper under double-blind review

ABSTRACT

Finite-step sampling from generative ODEs depends on where a solver evaluates the frozen vector field. We present Generative Inference Clock Optimization (GICO), an offline policy-optimization method for utility-specific allocation with the generator and solver rule fixed. Each clock is a normalized flow-time density whose inverse CDF yields a grid at the requested evaluation budget. A surrogate fitted to completed-solve utilities weights reference densities to anchor a policy conditioned on context, solver, and budget, and guides regularized refinement without further generator evaluations or gradients. An exact finite-budget adjoint identity explains utility sensitivity; smooth asymptotics yield an allocation surrogate. Experiments cover temporal sequences, molecular 3D coordinates, and images. Temporal errors decrease by up to 10.0% and Kabsch RMSD by 5.4–7.9% relative to uniform grids. Deterministic GICO attains the lowest mean FID among evaluated methods in 10/12 low-resolution image settings; its stochastic extension leads in the remaining two. FID reductions relative to fixed-reference envelopes reach 30.9%. With the full observation budget on frozen SANA, GICO policies outperform observation-matched Bayesian optimization and PPO-style clock policies on ImageReward and VQAScore at all tested NFEs. Held-out evaluations and ablations further support generalization across contexts and inference budgets without refitting.

1 INTRODUCTION

Continuous-time generative models transport a source distribution toward a data distribution using learned dynamics. Diffusion and score-based models derive sampling dynamics from a prescribed corruption process; continuous normalizing flows, flow matching, and rectified flow learn transport fields directly or along chosen paths (Chen et al., 2018; Grathwohl et al., 2019; Ho et al., 2020; Song et al., 2021b; Lipman et al., 2023; Liu et al., 2023). Smoothly increasing time changes with fixed endpoints preserve the exact ODE endpoint. Finite-step integration, however, generally breaks this invariance: the numerical endpoint depends on where the solver evaluates the frozen field.

For a generic generative ODE, we ask: *with the vector field, solver, and realized NFE fixed, how should field evaluations be allocated to improve terminal utility?* Probability paths and couplings can ease integration (Lipman et al., 2023; Liu et al., 2023; Tong et al., 2024), but do not by themselves determine the finite-budget grid best suited to a chosen utility. General transport fields also need not inherit a prescribed diffusion noise coordinate (Appx. A.6). The utility-aware allocation must therefore account for how finite-step numerical errors affect the quality measure evaluated on the completed samples. Existing methods optimize direct sampling grids (Sabour et al., 2024; Tong et al., 2025) or learn reward-driven instance-adaptive schedules (Ye et al., 2025; Ding et al., 2025; Yu et al., 2026); continuous trajectory transformations can support transfer across NFEs (Min et al., 2026). We explore offline contextual policy optimization: learning a reusable allocation policy from fixed completed-solve utilities while keeping the field and solver rule fixed. This requires a shared clock representation and a learning procedure that converts finite reference-clock evidence into conditional allocations across contexts and NFE budgets, without per-setting optimization at deployment.

Generative Inference Clock Optimization (GICO) addresses these issues by predicting a normalized flow-time density. Inverse-CDF decoding separates allocation from the number of solver intervals, while continuous NFE conditioning lets the density adjust to the requested budget. Together, they permit test-time deployment at unseen NFEs and contexts without refitting. A learned utility surrogate

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

weights reference densities to anchor the conditional policy and guides its regularized refinement. The resulting policy conditions on context, solver, and budget to form a grid without surrogate scoring or extra generator evaluations at deployment.

Our analysis separates the learning target from the allocation mechanism. A KL-regularized utility tilt defines a target law over reference clocks, whose forward-KL centroid yields the density anchor. Separately, an exact finite-budget adjoint identity relates full solver defects to terminal-utility error; its smooth asymptotic specialization gives an interpretable allocation surrogate, which motivates utility-based supervision rather than a field-only numerical error proxy. GICO uses measured finite-step utility without estimating adjoints or defect profiles. These characterize the learning construction and conditions for improvement, not unconditional guarantees for a fitted policy.

The contributions are threefold: (i) an offline policy-optimization approach that amortizes utility-specific allocation through a normalized density representation shared across settings; (ii) a finite-budget defect analysis and asymptotic allocation rule, with explicit reference weights and density anchors; and (iii) matched-NFE experiments across temporal, molecular, and image tasks, with representation, context, and held-out-budget controls. The deterministic policy deployment takes approximately 0.45–13.56 ms at batch size 64 across tasks (Appx. F.4). Anonymized code is available: `https://anonymous.4open.science/r/GenODE-Inference-6A5B`.

2 Related Work

Continuous-time generative models. Neural ODEs define continuous-depth dynamics, and continuous normalizing flows use invertible ODE transport and the instantaneous change-of-variables formula for likelihood modeling (Chen et al., 2018; Grathwohl et al., 2019). Diffusion and score-based models reverse a prescribed corruption process; with exact scores, their probability-flow ODEs share marginals with the stochastic dynamics (Ho et al., 2020; Song et al., 2021b). Flow matching, rectified flow, stochastic interpolants, and conditional flow matching train velocity fields along prescribed probability paths or couplings (Lipman et al., 2023; Liu et al., 2023; Albergo et al., 2025; Tong et al., 2024). These choices shape the trajectory geometry and integration difficulty, but local numerical fidelity and downstream generation utility are distinct objectives. GICO complements field and path design by changing where a frozen field is evaluated, not how its transport is learned.

Finite-step sampling and acceleration. DDIM, DPM-Solver, DEIS, EDM, and UniPC accelerate sampling through solver design (Song et al., 2021a; Lu et al., 2022; Zhang & Chen, 2023; Karras et al., 2022; Zhao et al., 2023); realized NFE distinguishes field-evaluation cost from step count. AYS, GITS, and OTS optimize discretization grids (Sabour et al., 2024; Chen et al., 2024; Xue et al., 2024). Bespoke Solvers learn a solver tailored to a pretrained generative flow, while Differentiable Solver Search jointly searches timesteps and solver coefficients (Shaul et al., 2024; Wang et al., 2025). LD3 and S4S use teacher-output matching, with S4S also learning solver coefficients (Tong et al., 2025; Frankel et al., 2025). BézierFlow learns continuous trajectory transformations that transfer across NFEs (Min et al., 2026). Score-Optimal, Entropic Time, and SharpEuler derive field-, information-, or acceleration-based scheduling criteria (Williams et al., 2024; Stancevic et al., 2025; Gupta et al., 2026). TPDM and RaSS adapt solver decisions during generation (Ye et al., 2025; Ding et al., 2025), while the James–Stein policy scheduler learns an instance-level interval policy by policy gradients (Yu et al., 2026). GICO also relates to weighted policy fitting (Peng et al., 2019) and offline optimization through learned objective models (Trabucco et al., 2021). Its utility-weighted anchors and surrogate refinement act in the node-count-independent density space, amortizing allocation across contexts, solvers, and budgets while keeping the field and solver rule fixed. The matched-NFE comparisons retain method-specific supervision, documented in Appx. B.6.

3 Preliminaries: Inference Clocks and Utility Tilting

3.1 Inference clocks as flow-time densities

Let $\boldsymbol{c}$ denote the context and $\boldsymbol{x}_0 \sim p_0(\cdot \mid \boldsymbol{c})$. We use column vectors and the Jacobian convention $[\boldsymbol{J}_{\boldsymbol{x}}\boldsymbol{f}]_{ij} = \partial f_i/\partial x_j$. The frozen generative ODE is defined by

$$\frac{\mathrm{d}\boldsymbol{x}_t}{\mathrm{d}t} = \boldsymbol{v}_\theta(\boldsymbol{x}_t, t; \boldsymbol{c}), \qquad t \in [0,1]. \tag{1}$$

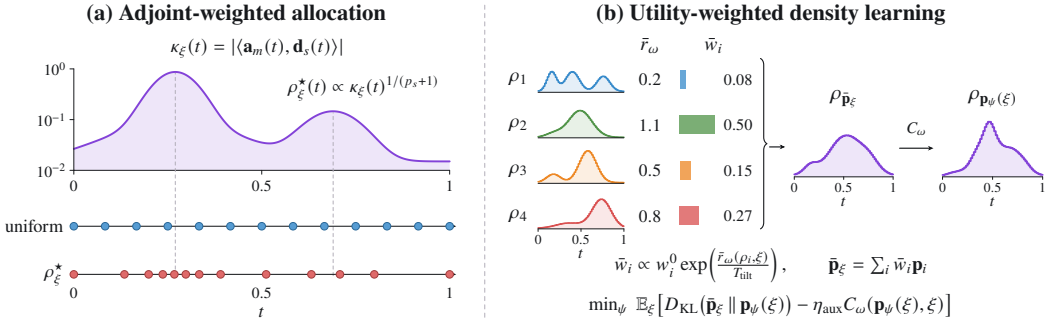

Figure 1: Utility-specific clock allocation: analysis and learning. **(a)** Adjoint-weighted asymptotics explain how terminal utility shapes the preferred allocation of the surrogate (Sec. 5). **(b)** A fitted utility surrogate weights reference densities to construct an anchor and guide regularized refinement of a conditional clock policy (Sec. 4).

An *inference clock* allocates solver evaluations over flow time t, represented as a positive normalized density

$$\mathcal{R} = \left\{ \rho \in L^1([0,1]) : \rho(t) > 0 \text{ a.e.}, \int_0^1 \rho(t)\,\mathrm{d}t = 1 \right\}. \tag{2}$$

For $\rho \in \mathcal{R}$, define its cumulative mass and inverse

$$F_\rho(t) = \int_0^t \rho(a)\,\mathrm{d}a, \qquad \phi_\rho(\lambda) = \inf\{t \in [0,1] : F_\rho(t) \geq \lambda\}, \qquad \lambda \in [0,1]. \tag{3}$$

Here, λ is clock time and $t = \phi_\rho(\lambda)$ is flow time. Positivity almost everywhere makes F_ρ continuous and strictly increasing. Let s be the solver, b its realized NFE, and $K = K_s(b)$ its macro-step count. The boundaries are

$$\tau_k^{\rho,s,b} = \phi_\rho\left(\frac{k}{K}\right), \qquad k = 0, \ldots, K. \tag{4}$$

Internal field evaluations follow the fixed solver rule. Each interval carries density mass $1/K$, so greater density allocates finer resolution. Write $\tau_k^\rho := \tau_k^{\rho,s,b}$ and $\Delta t_k^\rho := \tau_{k+1}^\rho - \tau_k^\rho$.

For a given partition of flow time into B bins, positive probability masses $\boldsymbol{p} \in \Delta_B^\circ$ define a piecewise-constant density $\rho_{\boldsymbol{p}}$, constant within each bin. Here, Δ_B is the probability simplex. Its dimension is independent of K, so prediction, averaging, and regularization share the same allocation space as the node count changes. Appx. B.1 provides the numerical realization.

Smooth re-clocking preserves the exact ODE endpoint. At finite NFE, however, applying the solver on the induced t-grid need not coincide with integrating the transformed ODE uniformly in clock time. Here, clocks specify node placement for the original field and solver. Appx. A.1–A.3 derive the time change, an Euler example of finite-step dependence, and the node-density interpretation.

3.2 Finite-step utility and its tilted clock law

Let $\xi = (m, s, b, \boldsymbol{c})$ collect the task, solver, budget, and context. The utility specification m fixes the metrics, transformations, and scalarization weights throughout each learning problem. A clock induces the numerical endpoint

$$\widehat{\boldsymbol{x}}_1^\rho = \mathrm{Solve}_s\left(\boldsymbol{v}_\theta, \boldsymbol{x}_0, \boldsymbol{c}, \boldsymbol{\tau}^{\rho,s,b}\right). \tag{5}$$

For a fixed higher-is-better utility U_m, possibly evaluated jointly on a batch of endpoints, define

$$S_\xi(\rho) = \mathbb{E}[U_m(\widehat{\boldsymbol{x}}_1^\rho, \boldsymbol{c}) \mid \rho, \xi]. \tag{6}$$

Selecting a whole clock at ξ is a contextual-bandit decision with expected utility $S_\xi(\rho)$, averaging source and evaluation randomness. A reference law $q_0(\mathrm{d}\rho \mid \xi)$ on a measurable feasible family $\mathcal{R}_\mathrm{f} \subseteq \mathcal{R}$ gives the KL-regularized objective:

$$\max_{\pi \ll q_0} \left\{ \mathbb{E}_{\rho\sim\pi}[S_\xi(\rho)] - \frac{1}{\beta} D_{\text{KL}}(\pi \,\|\, q_0) \right\}. \tag{7}$$

For $\beta > 0$, $0 < Z_\beta(\xi) < \infty$, and finite expected utility and KL of the tilted law, the unique solution is

$$Z_\beta(\xi) = \int_{\mathcal{R}_\mathrm{f}} e^{\beta S_\xi(\widetilde{\rho})} q_0(\mathrm{d}\widetilde{\rho} \mid \xi), \qquad \pi^\star_\beta(\mathrm{d}\rho \mid \xi) = \frac{e^{\beta S_\xi(\rho)}}{Z_\beta(\xi)} q_0(\mathrm{d}\rho \mid \xi). \tag{8}$$

This Gibbs variational characterization (Appx. A.7) turns clock preferences into a regularized target distribution. The reference law determines accessible clocks; β controls how strongly utility changes their relative weights. A density ρ allocates evaluations over time, whereas $\pi^\star_\beta$ assigns probability to entire clocks. The target law in Eq. (8) applies directly to finite-budget utility. Sec. 4 constructs learning targets using its finite-reference specialization (Alg. 1), while Sec. 5 analyzes the source of utility-dependent preferences.

4 Learning Inference Clocks from Completed Solves

Algorithm 1 Generative Inference Clock Optimization (GICO)

Require: Frozen field $\boldsymbol{v}_\theta$, settings Ξ, reference families $\{\mathcal{C}_\xi\}$, and utility specification m.
Ensure: Deterministic density policy $\boldsymbol{p}_\psi$ or stochastic clock policy P_ψ

1: Collect reference-clock utilities and paired uniform measurements before fitting (Eq. (9)).
2: Fit the utility surrogate $\boldsymbol{r}_\omega$ by ranking and regression (Eq. (10)).
3: Construct utility-weighted anchors (Eq. (11); stochastic: Appx. A.11).
4: Optimize the anchored policy objective (Eq. (12); stochastic: Appx. A.11).
5: For a requested setting, predict or sample a density and decode its grid by Eq. (4).
6: **return** the conditional clock policy.

4.1 Measured utilities and the utility surrogate

For each ξ, let $\mathcal{C}_\xi = (\rho_i)_{i=1}^{n_\xi}$ contain reference clocks with positive base weights $w_i^0(\xi)$ summing to one. They define the empirical design law $q_{0,n_\xi} = \sum_i w_i^0(\xi)\delta_{\rho_i}$. For collected references, source draws $\mathcal{D}_\xi = \{\boldsymbol{x}_0^{(r)}\}_{r=1}^{N_\xi}$ shared with the uniform anchor give

$$\widehat{\boldsymbol{u}}_i(\xi) = \widehat{\boldsymbol{U}}_{m,\mathcal{D}_\xi}\left(\left\{\mathrm{Solve}_s(\boldsymbol{v}_\theta, \boldsymbol{x}_0^{(r)}, \boldsymbol{c}, \boldsymbol{\tau}^{\rho_i,s,b})\right\}_{r=1}^{N_\xi}, \boldsymbol{c}\right), \qquad \widehat{S}_i(\xi) = \widetilde{\boldsymbol{\alpha}}_m^\top \widehat{\boldsymbol{u}}_i(\xi). \tag{9}$$

Here $\widehat{\boldsymbol{U}}_{m,\mathcal{D}_\xi}$ averages samplewise components over draws or evaluates batchwise components jointly. The nonnegative component weights are fixed by the utility specification. The target can thus express sample quality or a distributional criterion, rather than only fidelity to a high-accuracy numerical endpoint or rollout. The metric estimators and scalarizations are specified in Appx. D.2–D.3.

The *utility surrogate* $\boldsymbol{r}_\omega(\rho, \xi)$ predicts utility components, with scalar score $\bar{r}_\omega(\rho, \xi) = \widetilde{\boldsymbol{\alpha}}_m^\top \boldsymbol{r}_\omega(\rho, \xi)$. It is fitted to the measured reference utilities using

$$\min_\omega \ \mathcal{L}_\mathrm{surr}(\omega) = \mathcal{L}_\mathrm{rank}(\omega) + \lambda_\mathrm{reg}\mathcal{L}_\mathrm{reg}(\omega). \tag{10}$$

Ranking learns the within-setting preferences; regression calibrates score gaps for Gibbs weighting, whose effective temperature depends on score scale. The collected utilities remain fixed throughout surrogate and policy fitting. The surrogate predicts clock utilities rather than target samples. Objective and fitting details are in Appx. B.3 and C.2.

4.2 Utility-weighted density targets

Let $\boldsymbol{p}_i$ denote the bin masses of ρ_i. The surrogate scores every reference in $\mathcal{C}_\xi$, including unmeasured context–clock pairs. Applying the finite-reference tilt to these predictions with $\beta = 1/T_\mathrm{tilt}$ gives

$$\begin{aligned} \bar{w}_i(\xi) &= \frac{w_i^0(\xi) e^{\bar{r}_\omega(\rho_i,\xi)/T_\mathrm{tilt}}}{\sum_{j=1}^{n_\xi} w_j^0(\xi) e^{\bar{r}_\omega(\rho_j,\xi)/T_\mathrm{tilt}}}, \\ \bar{\boldsymbol{p}}_\xi &= \sum_{i=1}^{n_\xi} \bar{w}_i(\xi)\boldsymbol{p}_i = \arg\min_{\boldsymbol{p}\in\Delta_B} \sum_{i=1}^{n_\xi} \bar{w}_i(\xi) D_\mathrm{KL}(\boldsymbol{p}_i \,\|\, \boldsymbol{p}). \end{aligned} \tag{11}$$

Eq. (11) performs two operations: the weights define the finite tilted law $\widehat{\pi}_\omega = \sum_i \bar{w}_i \delta_{\rho_i}$, and its forward-KL centroid yields the density anchor (Appx. A.10). Linear reconstruction gives $\rho_{\bar{\boldsymbol{p}}_\xi} = \sum_i \bar{w}_i \rho_{\boldsymbol{p}_i}$, so the anchor remains a normalized density in the same family. It thus averages reference allocations in a common, node-count-independent density space, rather than selecting one clock or averaging individual grid nodes.

4.3 Anchored conditional policy optimization

The deterministic policy predicts $\boldsymbol{p}_\psi(\xi) \in \Delta_B^\circ$. Let $C_\omega(\boldsymbol{p}, \xi)$ be a normalized utility score derived from the frozen utility surrogate (Appx. B.4). For $\eta_{\mathrm{aux}} \geq 0$, we minimize

$$\min_\psi \ \mathcal{L}_{\mathrm{pol}}(\psi) = \mathbb{E}_\xi[D_{\mathrm{KL}}(\bar{\boldsymbol{p}}_\xi \,\|\, \boldsymbol{p}_\psi(\xi)) - \eta_{\mathrm{aux}} C_\omega(\boldsymbol{p}_\psi(\xi), \xi)]\,. \tag{12}$$

This KL acts on allocation densities, not laws over clocks, and equals the utility-weighted forward KL from reference allocations up to a constant. At $\eta_{\mathrm{aux}} = 0$, the per-setting optimum is the centroid. For $\eta_{\mathrm{aux}} > 0$, a per-setting solution with objective value no greater than the anchor's can depart from it only if the surrogate-utility gain offsets the KL penalty. This permits refinement beyond the scored clocks while regularizing departures from the anchor, with a local natural-gradient interpretation under the categorical Fisher metric (Appx. A.10). Thus, the shared policy amortizes this optimization across contexts, solvers, and budgets; Eq. (4) then converts the predicted density into the grid at the requested solver budget. Parameterization and fitting details are in Appx. B–C.

Stochastic extension. The same utility-weighted reference law supports a stochastic policy: weighted autoregressive likelihood fits a distribution over clock representations, while expected surrogate utility guides refinement over its samples. A clock is sampled before ODE integration. Appx. A.9–A.11 relate the population likelihood target to the Doob transform on clock-construction paths and define the empirical target and objective.

5 Allocation Analysis and Learning Interpretation

Fix a source draw and $(m, s, \boldsymbol{c})$. For smooth one-step maps and utility u_m, the exact finite-budget utility error is $-\sum_k \langle \bar{\boldsymbol{a}}_{k+1}, \boldsymbol{\delta}^\rho_{s,k} \rangle$, using full local defects and averaged discrete adjoints (Appx. A.4). This identity retains nonlinear propagation without a small-step approximation. Under the stated smoothness and stability assumptions, an order-p_s solver's principal defect $\boldsymbol{d}_s(\boldsymbol{x}_t, t; \boldsymbol{c})(\Delta t)^{p_s+1}$ and continuous adjoint $\boldsymbol{a}_m(t) = [\boldsymbol{J}_{\boldsymbol{x}} \Phi_{t,1}(\boldsymbol{x}_t)]^\top \nabla_{\boldsymbol{x}} u_m(\boldsymbol{x}_1, \boldsymbol{c})$, with exact flow map Φ, give, as $K = K_s(b_K) \to \infty$,

$$u_m(\widehat{\boldsymbol{x}}^\rho_1, \boldsymbol{c}) - u_m(\boldsymbol{x}_1, \boldsymbol{c}) = -\sum_{k=0}^{K-1} \left\langle \boldsymbol{a}_m(\tau^\rho_{k+1}), \boldsymbol{d}_s(\boldsymbol{x}_{\tau^\rho_k}, \tau^\rho_k; \boldsymbol{c}) \right\rangle (\Delta t^\rho_k)^{p_s+1} + o(K^{-p_s}), \tag{13}$$

uniformly over common density bounds $0 < \rho_{\min} \leq \rho \leq \rho_{\max} < \infty$. Both forms weight defects by terminal sensitivity rather than their norms, and retain signed cancellation. Appx. A.4 also bounds the finite-budget linearization remainder and gives a sufficient condition for allocation improvement; large numerical amplification can make this bound loose.

To isolate the allocation effect, consider a nonnegative profile $\kappa_\xi(t)$ that ignores signed cancellation and is treated as clock-independent. Inverse-CDF placement gives approximate local widths $[K\rho(t)]^{-1}$ and interval count $K\rho(t)\,\mathrm{d}t$, yielding the allocation surrogate

$$\mathcal{A}^{\mathrm{disc}}_{\xi,K}(\rho) = \sum_{k=0}^{K-1} \kappa_\xi(\tau^\rho_k)(\Delta t^\rho_k)^{p_s+1} \ \approx \ \mathcal{A}_\xi(\rho) := K^{-p_s} \int_0^1 \kappa_\xi(t)\rho(t)^{-p_s}\,\mathrm{d}t. \tag{14}$$

For smooth utilities, an average or envelope of $|\langle \boldsymbol{a}_m(t), \boldsymbol{d}_s(\boldsymbol{x}_t, t; \boldsymbol{c})\rangle|$ motivates κ_ξ. This clock-independent profile is a surrogate, not an exact low-NFE or nonsmooth-utility representation. If $\kappa_\xi > 0$ almost everywhere and $\kappa_\xi^{1/(p_s+1)} \in L^1$, strict convexity gives the unique unconstrained surrogate minimizer

$$\rho^\star_\xi(t) = \frac{\kappa_\xi(t)^{1/(p_s+1)}}{\int_0^1 \kappa_\xi(a)^{1/(p_s+1)}\,\mathrm{d}a}. \tag{15}$$

This assigns denser resolution where local defects have larger propagated effects on terminal utility (Appx. A.5). For a fixed profile, K^{-p_s} changes only the surrogate's scale, not its optimum. Finite-budget preferences can differ beyond this leading-order model. GICO thereby shares an allocation representation across budgets but allows the predicted density to depend on NFE, using completed-solve utility as supervision.

A synthetic example. In an exactly solvable two-dimensional ODE, analytical allocations reduce Euler terminal-utility error by $14.77\%/43.47\%$ for two utilities at NFE 2 and $20.05\%/49.10\%$ at NFE 4. Full-defect identities remain exact where leading-order predictions are inaccurate (Appx. A.5).

Connection to learning. Finite-budget defect propagation motivates terminal-utility supervision; Eq. (15) characterizes only the leading-order allocation (Fig. 1a). GICO estimates neither adjoints nor this profile. Measured finite-step utilities instead induce the weights and anchor in Eq. (11) (Fig. 1b), followed by Eq. (12) refinement. Appx. A.10 gives a calibration-dependent sufficient condition for true-utility improvement over the anchor, not an unconditional guarantee for the fitted policy.

6 Experiments

Protocol and setup. Experiments and comparisons fix the generator, context encoder, solver, realized NFE, and source seeds; test data are excluded from fitting and selection. Temporal sequences (solar, traffic, weather (Godahewa et al., 2021)) and molecular relaxation coordinates (sets 1–3; Appx. D.1) use four solvers; CIFAR-10 (Krizhevsky, 2009) and EDM-prepared ImageNet-64 (Karras et al., 2022) use Euler. References include uniform, Late-p, transferred AYS/GITS/OTS, a FlowTS-inspired t-power clock, and reversals (Sabour et al., 2024; Chen et al., 2024; Xue et al., 2024; Hu et al., 2024). Appx. B.2, D.1, and D.3 specify the mappings and evaluation. COCO (Lin et al., 2014) uses frozen SANA/FlowEuler and SD1.5/iPNDM-2 (Xie et al., 2025; Rombach et al., 2022), comparing deterministic global and prompt-conditioned GICO (Appx. G.1–G.3).

6.1 Low-NFE Downstream Performance

Temporal and 3D. GICO improves over fixed-reference envelopes in 20/24 temporal and all 24 molecular aggregate cells (Tabs. 1–2), using four-solver means at the final 20k field checkpoint. C, M, and K denote rCRPS, rMASE, and rKabsch relative to uniform (lower is better); U is dynamics utility U_{comp}, excluding Kabsch (higher is better). Envelopes select within each family/solver and then select a family after solver averaging, rather than using a solver-wise oracle. Traffic and weather improve both temporal metrics at every tested NFE; across the 20 winning temporal cells, errors decrease by 1.1–10.0% relative to uniform. Molecular Kabsch error decreases by 5.4–7.9%, with positive dynamics utility in every aggregate setting. These gains cover coordinate fidelity and the velocity/acceleration terms included in the fitted utility (Eq. (9); Appx. D.2), rather than coordinate error alone. Appendix SDs report solver spread (Appx. E.1) and matched comparisons (Appx. F.5).

Low-resolution images. Deterministic GICO lowers mean FID versus the fixed-reference envelope in all 12 settings and leads all evaluated methods in 10/12 (Tab. 3); stochastic GICO leads CIFAR-10 RF++ at NFE 8 and ImageNet-64 RF++ at NFE 4. ImageNet-64 EDM-VE→RF FID at NFE 8 falls from 21.820 to 15.077, a 30.9% reduction. KID supplies supervision, whereas FID, IS, precision, and recall are evaluation-only metrics, so the FID gains extend beyond the fitting metric. Across temporal, molecular, and image studies, deterministic GICO improves 85/96 aggregate mean cells over fixed-reference envelopes: 20/24, 24/24, and 41/48. Full tables are in Appx. E.2–E.4.

Frozen text-to-image models. Tab. 4 reports ImageReward (IR) and VQAScore (VQA) on 512 held-out prompts, testing generalization of completed-solve reward supervision across disjoint contexts. At SANA NFE 4, GICO-global gains 0.6292 IR over native and 0.1044 over selected fixed, with paired 95% intervals [0.5747, 0.6842] and [0.0711, 0.1386]; at NFE 8, gains over selected fixed reach 0.0280 IR and 0.0051 VQA, both with positive intervals (Appx. G.5). At the full 6,400-observation budget, GICO-global exceeds BO and GICO-prompt exceeds PPO-style clock PG on both SANA rewards at every tested NFE, again with positive paired intervals; at NFE 4, the latter gains are 0.5264 IR and 0.0789 VQA, and GICO-prompt also leads at 1,600/3,200 observations (Fig. 9). Against our separate JS-REINFORCE reconstruction, GICO leads in five of six metric–NFE cells (Appx. G.8).

Table 1: Temporal: GICO / reference envelope.

Dataset	NFE 4	NFE 6	NFE 8	NFE 10
Solar/C	**.961**/.998	**.958**/.993	1.004/**.987**	1.007/**.990**
Solar/M	**.989**/1.000	**.982**/.997	1.005/**.993**	1.002/**.986**
Traffic/C	**.980**/.983	**.963**/.971	**.958**/.968	**.958**/.971
Traffic/M	**.946**/.961	**.928**/.940	**.910**/.924	**.908**/.914
Weather/C	**.934**/.950	**.925**/.934	**.914**/.929	**.914**/.934
Weather/M	**.925**/.932	**.917**/.919	**.901**/.921	**.900**/.919

Table 2: 3D: GICO / reference envelope.

Dataset	NFE 4	NFE 6	NFE 8	NFE 10
Set 1/K	**.945**/.991	**.935**/.989	**.926**/.989	**.923**/.991
Set 1/U	**.064**/.002	**.078**/.001	**.089**/.000	**.093**/.000
Set 2/K	**.946**/.984	**.937**/.983	**.929**/.986	**.927**/.985
Set 2/U	**.057**/.003	**.069**/.003	**.079**/.000	**.082**/.001
Set 3/K	**.946**/.991	**.936**/.986	**.923**/.983	**.921**/.981
Set 3/U	**.067**/.003	**.079**/.001	**.087**/.000	**.090**/.001

Table 3: Image FID50k ↓ (Euler, five seeds). Bold/underline: best/second-best.

Dataset / backbone	NFE	Reference envelope	SharpEuler	GICO-det.	GICO-stoch.
CIFAR-10 RF++	2 (seen)	2.466	<u>2.438</u>	**2.402**	2.478
	4 (unseen)	2.290	<u>2.225</u>	**2.218**	2.266
	8 (seen)	2.215	<u>2.187</u>	2.210	**2.002**
CIFAR-10 EDM-VE→RF	2 (seen)	128.943	<u>124.802</u>	**121.247**	130.765
	4 (unseen)	41.955	<u>39.913</u>	**39.542**	49.015
	8 (seen)	15.450	<u>14.568</u>	**14.244**	18.303
ImageNet-64 RF++	2 (seen)	<u>3.655</u>	4.531	**3.334**	3.661
	4 (unseen)	3.324	4.187	<u>3.061</u>	**2.986**
	8 (seen)	3.200	4.070	**2.930**	<u>3.192</u>
ImageNet-64 EDM-VE→RF	2 (seen)	174.098	186.759	**152.510**	<u>172.350</u>
	4 (unseen)	<u>59.235</u>	89.421	**43.718**	74.732
	8 (seen)	21.820	23.047	**15.077**	<u>21.507</u>

Table 4: COCO text-to-image at full observation budget (IR and VQA ↑).

	NFE 4		NFE 6		NFE 8	
Method	IR	VQA	IR	VQA	IR	VQA
SANA-600M-512px						
Uniform	−0.6799	0.6391	0.2441	0.7972	0.6140	0.8432
Native	−0.1864	0.7191	0.7347	0.8488	0.8830	0.8696
Selected fixed	0.3384	0.8092	0.7760	0.8502	0.8603	0.8648
Bayesian optimization	0.3382	0.7960	0.5174	0.8275	0.6140	0.8432
PPO-style clock PG	−0.1028	0.7310	0.5328	0.8250	0.6546	0.8434
GICO-global	0.4429	0.8167	0.7531	0.8464	0.8883	0.8699
GICO-prompt	0.4236	0.8099	0.7700	0.8534	0.8831	0.8682
SD1.5 EMA						
Native / uniform	−0.4155	0.7235	−0.0153	0.7688	0.1028	0.7819
Selected fixed	−0.4155	0.7235	−0.0510	0.7565	0.1238	0.7812
LD3	−0.4513	0.7228	−0.1741	0.7429	0.0481	0.7730
GICO-global	−0.4117	0.7220	−0.0243	0.7666	0.1837	0.7711
GICO-prompt	−0.4212	0.7175	−0.0478	0.7606	0.1705	0.7732

6.2 Budget Transfer, Design Controls, and Discussion

Held-out NFEs. Transfer tests the separation between density prediction and node count in Secs. 3–4. Target NFEs are excluded from fitting, normalization, and selection (Appx. C.3). The temporal/3D policies fit at 4/8 NFEs and transfer to 6 (interpolation) and 10 (extrapolation); low-resolution images transfer 2/8-to-4. The deterministic GICO improves 36/40 held-out mean cells over fixed-reference envelopes (22/24 temporal/3D; 14/16 image). On frozen BézierFlow, 4/8/10-to-6 transfer reduces FID by 9.0%, showing that learned allocation can transfer on top of an existing trajectory transformation (Appx. E.6). COCO provides another zero-shot test: policies fitted at 4/8 deploy at NFE 6 without refitting, with networks and normalizers frozen. At full transfer budget, SANA global/prompt policies exceed their direct-fit counterparts by +0.0215/+0.0248 IR, while VQA changes remain near zero; transfer has twice the observation allowance of direct fitting (Appx. G.6).

Structure and controls. The direct-grid ablation replaces density decoding with direct node prediction; the context-free control removes the context while retaining solver and NFE inputs. Under



Figure 2: Matched COCO examples: **A**, SANA at NFE 4; **B**, SD1.5 at NFE 6. Prompts and noise are shared across methods. Score-independent selection and the extended gallery: Appx. G.10.

matched supervision and optimization, their metric ratios to GICO are 1.300 and 1.187, respectively, and 1.364 when combined (Appx. F.2). These ablations support the shared density representation and conditional allocation of Sec. 4. Learned densities also vary across settings (Fig. 3), consistent with Sec. 5: propagated utility sensitivity can favor different allocations across tasks, contexts, and solvers. Stochastic image policies retain within-context clock diversity while their mean-density total variation distances from deterministic GICO peak at 0.1391/0.1225 on CIFAR-10/ImageNet-64 (Appx. F.3). Across the extended temporal/3D diagnostic, aggregate effects remain favorable from 4k to 20k frozen-field checkpoints and NFEs 4–20, although their magnitude varies non-monotonically with field maturity and budget (Appx. F.1). In the SANA NFE 6 controls, GICO-prompt exceeds surrogate top-1 by 0.0481 IR with no resolved VQA difference; IR-only weighting increases IR by 0.0236 without a resolved VQA change, while neither metric resolves a difference between 12-clock support and the 25-clock baseline (Appx. G.7). These controls distinguish density learning from discrete reference selection and show that reference count alone does not determine performance.

Scope and discussion. Gains are not universal: Solar errors rise by 0.2–0.7% at NFEs 8/10, and contextwise distributions include negative tails (Fig. 4). Off-hull densities can also decode to grids reproducible by reference mixtures (Appx. G.11). Fixed or native schedules remain competitive: at NFE 6, selected fixed exceeds GICO-global on SANA and native leads on SD1.5, while SD1.5 at NFE 8 trades higher IR for lower VQA. Utility also matters: at SANA NFE 6, VQA-only weighting lowers both IR and VQA relative to balanced GICO-prompt, and several COCO baseline and transfer intervals span zero (Appx. G.5–G.7). GICO fits fixed evidence, whereas BO and clock PG evaluate adaptive proposals; comparisons match utility-observation allowances, not fitting computation or supervision. LD3 matches inference NFE only and uses LPIPS teacher targets (Appx. G.3); the separate JS-REINFORCE reconstruction has a higher mean SANA VQA at NFE 4 (Appx. G.8).

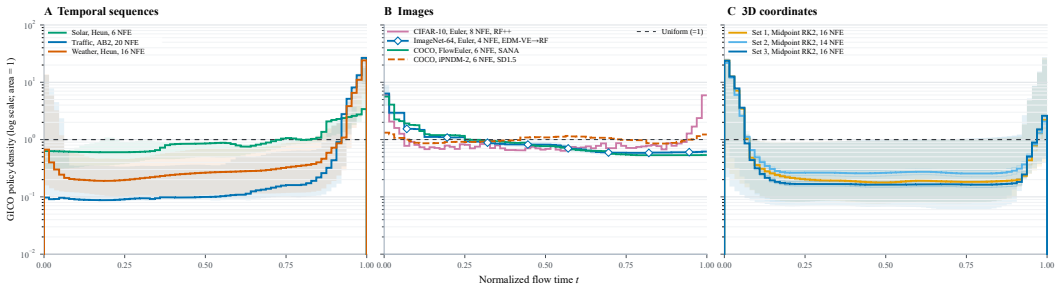

Figure 3: Learned normalized flow-time densities (log scale), not direct estimates of κ_ξ; shading shows context-level variation. Values above one indicate denser allocation and vice versa. Additional learned deterministic and stochastic densities are shown in Fig. 6.

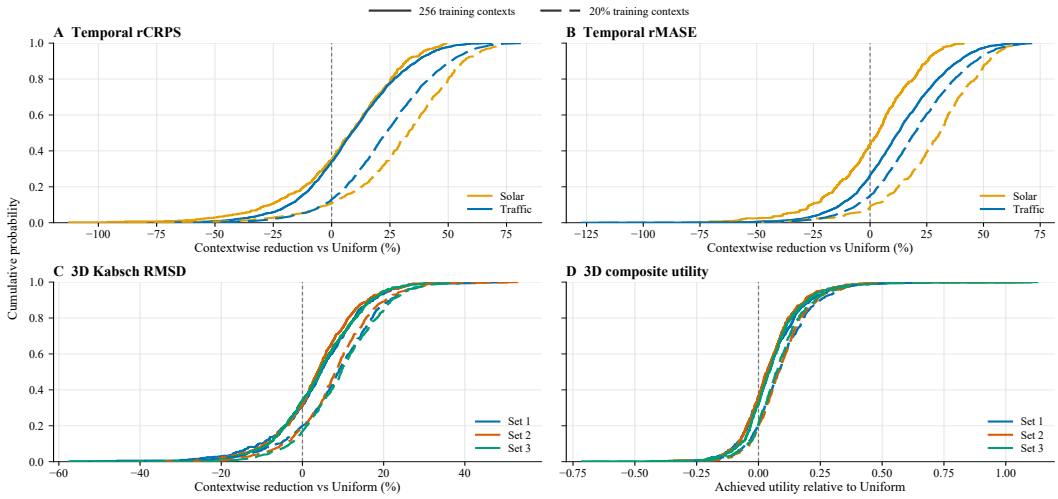

Figure 4: Context-wise improvement over uniform with the stochastic policy. Positive is favorable; solid/dashed curves: 256 contexts/20% train set contexts. Negative tails show individual regressions.

Composition with frozen BézierFlow is likewise budget-dependent: re-clocking lowers FID at NFEs 4/6/8 but raises it by 16.0% at NFE 10 (Appx. E.6).

7 Conclusion

GICO formulates finite-step inference scheduling as offline contextual policy optimization over normalized flow-time densities. Utility-weighted anchors and surrogate refinement amortize allocation across contexts, solvers, and NFE budgets with the field and update rule fixed. An exact finite-budget identity relates full defects to smooth-utility errors; its asymptotic allocation surrogate motivates completed-solve supervision rather than defect norms alone.

Across temporal sequences, molecular coordinates, and images, GICO improves most settings over fixed-reference envelopes and transfers to held-out NFEs. KID-supervised image policies improve evaluation-only FID and IS, while SANA experiments show gains over observation-matched BO and PPO-style clock policies at full budget. Density and context ablations, together with surrogate-top-1 controls, support amortized clock optimization beyond discrete node prediction or reference selection.

Nevertheless, performance can depend on utility specification, reference coverage, and surrogate calibration; aggregate gains need not extend to every context or budget. Finite-budget bounds can be loose when defects or propagation factors are large, and the power-law allocation remains asymptotic. Computable certificates and active acquisition of informative clock–utility pairs through stochastic policies are natural next steps.

AI USE STATEMENT

We used ChatGPT and Claude for manuscript polish and LaTeX file management, and ChatGPT for additional template, citation, and cross-reference checks. The authors are responsible for verifying all claims, proofs, citations, and experimental results.

ETHICS STATEMENT

This work optimizes inference-time node placement for frozen generative ODEs; it does not collect new data or retrain the underlying generators. The approach nevertheless inherits the risks of the datasets, pretrained models, utility functions, and applications with which it is used, including bias, privacy and licensing concerns, unsafe outputs, and misuse. Moreover, lower-cost sampling can enable beneficial deployment but can also increase the scale of harmful generation. Utility misspecification may favor aggregate metrics while obscuring failures for particular contexts or groups, consistent with the negative tails observed in our contextwise results. Responsible use therefore requires application-specific auditing, compliance with dataset licenses and consent or privacy requirements, appropriate safeguards, and human oversight.

REPRODUCIBILITY STATEMENT

Sections 3–5 specify the clock representation, utility target, implemented learning objectives, and allocation analysis; Sec. 6 gives the evaluation design. Appx A states the assumptions and gives the derivations supporting the theoretical claims. Appx B documents the clock representation, surrogate and policy objectives, numerical safeguards, and deployment. Architectures, optimization, checkpoint selection, and NFE conditioning are in Appx C; datasets, rewards, NFE splits, and matched evaluation are in Appx D. Appx G specifies COCO reward targets, observation budgets, baseline differences, paired intervals, and transfer controls. Experiments hold the generator, solver, realized NFE, conditioning information, and source seeds fixed across clock comparisons, and locked evaluation data are excluded from fitting and checkpoint selection. An anonymized code repository supporting replication is linked in the introduction. Text-to-image experiments use NVIDIA H100 GPUs; other experiments use A100 GPUs.

REFERENCES

Michael Albergo, Nicholas M. Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *Journal of Machine Learning Research*, 26(209):1–80, 2025. URL https://www.jmlr.org/papers/v26/23-1605.html.

Sebastian Ament, Samuel Daulton, David Eriksson, Maximilian Balandat, and Eytan Bakshy. Unexpected improvements to expected improvement for Bayesian optimization. In *Advances in Neural Information Processing Systems*, volume 36, pp. 20577–20612, 2023. doi: 10.52202/075280-0904. URL https://proceedings.neurips.cc/paper/2023/hash/419f72cbd568ad62183f8132a3605a2a-Abstract-Conference.html.

Maximilian Balandat, Brian Karrer, Daniel Jiang, Samuel Daulton, Ben Letham, Andrew G. Wilson, and Eytan Bakshy. BoTorch: A framework for efficient Monte-Carlo Bayesian optimization. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/f5b1b89d98b7286673128a5fb112cb9a-Abstract.html.

Mikołaj Bińkowski, Danica J. Sutherland, Michael Arbel, and Arthur Gretton. Demystifying MMD GANs. In *International Conference on Learning Representations*, 2018. URL https://arxiv.org/abs/1801.01401.

Kevin Black, Michael Janner, Yilun Du, Ilya Kostrikov, and Sergey Levine. Training diffusion models with reinforcement learning. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/14f75513f0f1ca01de1e826b52e6b840-Abstract-Conference.html.

Yang Cao and Linda Petzold. A posteriori error estimation and global error control for ordinary differential equations by the adjoint method. *SIAM Journal on Scientific Computing*, 26(2): 359–374, 2004. doi: 10.1137/S1064827503420969.

Defang Chen, Zhenyu Zhou, Can Wang, Chunhua Shen, and Siwei Lyu. On the trajectory regularity of ODE-based diffusion sampling. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 7905–7934. PMLR, 2024. URL https://proceedings.mlr.press/v235/chen24bm.html.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K. Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, volume 31, pp. 6571–6583, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/69386f6bb1dfed68692a24c8686939b9-Abstract.html.

Xinlei Chen, Hao Fang, Tsung-Yi Lin, Ramakrishna Vedantam, Saurabh Gupta, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO captions: Data collection and evaluation server. *arXiv preprint arXiv:1504.00325*, 2015. URL https://arxiv.org/abs/1504.00325.

Raphaël Chetrite and Hugo Touchette. Nonequilibrium Markov processes conditioned on large deviations. *Annales Henri Poincaré*, 16:2005–2057, 2015. doi: 10.1007/s00023-014-0375-8. URL https://link.springer.com/article/10.1007/s00023-014-0375-8.

Patryk Chrabaszcz, Ilya Loshchilov, and Frank Hutter. A downsampled variant of ImageNet as an alternative to the CIFAR datasets. *arXiv preprint arXiv:1707.08819*, 2017. URL https://arxiv.org/abs/1707.08819.

Xin Ding, Lei Yu, Xin Li, Zhijun Tu, Hanting Chen, Jie Hu, and Zhibo Chen. RaSS: Improving denoising diffusion samplers with reinforced active sampling scheduler. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12923–12933, 2025. URL https://openaccess.thecvf.com/content/CVPR2025/html/Ding_RaSS_Improving_Denoising_Diffusion_Samplers_with_Reinforced_Active_Sampling_Scheduler_CVPR_2025_paper.html.

J. L. Doob. Conditional Brownian motion and the boundary limits of harmonic functions. *Bulletin de la Société Mathématique de France*, 85:431–458, 1957. doi: 10.24033/bsmf.1494. URL https://www.numdam.org/articles/10.24033/bsmf.1494/.

Eric Frankel, Sitan Chen, Jerry Li, Pang Wei Koh, Lillian J. Ratliff, and Sewoong Oh. S4S: Solving for a fast diffusion model solver. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 17542–17574, 2025. URL https://proceedings.mlr.press/v267/frankel25a.html.

Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi: 10.1198/016214506000001437.

Rakshitha Godahewa, Christoph Bergmeir, Geoffrey I. Webb, Rob J. Hyndman, and Pablo Montero-Manso. Monash time series forecasting archive. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, volume 1, 2021. URL https://openreview.net/forum?id=wEc1mgAjU-.

Will Grathwohl, Ricky T. Q. Chen, Jesse Bettencourt, Ilya Sutskever, and David Duvenaud. FFJORD: Free-form continuous dynamics for scalable reversible generative models. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=rJxgknCcK7.

Aditi Gupta, Soon Hoe Lim, Annan Yu, and N. Benjamin Erichson. Sharpen your flow: Sharpness-aware sampling for flow matching. *arXiv preprint arXiv:2605.11547*, 2026. URL https://arxiv.org/abs/2605.11547.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/8a1d694707eb0fefe65871369074926d-Abstract.html.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022. URL https://arxiv.org/abs/2207.12598.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020. URL https://arxiv.org/abs/2006.11239.

Yang Hu, Xiao Wang, Zezhen Ding, Lirong Wu, Huatian Zhang, Stan Z. Li, Sheng Wang, Jiheng Zhang, Ziyun Li, and Tianlong Chen. FlowTS: Time series generation via rectified flow. *arXiv preprint arXiv:2411.07506*, 2024. URL https://arxiv.org/abs/2411.07506.

Peter J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35 (1):73–101, 1964. doi: 10.1214/aoms/1177703732.

Rob J. Hyndman and Anne B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, 2006. doi: 10.1016/j.ijforecast.2006.03.001.

W. Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, 32(5):922–923, 1976. doi: 10.1107/S0567739476001873.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 26565–26577, 2022. doi: 10.52202/068431-1926.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009. URL https://cave.cs.toronto.edu/kriz/learning-features-2009-TR.pdf.

Tuomas Kynkäänniemi, Tero Karras, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Improved precision and recall metric for assessing generative models. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/0234c510bc6d908b28c70ff313743079-Abstract.html.

Sangyun Lee, Zinan Lin, and Giulia Fanti. Improving the training of rectified flows. In *Advances in Neural Information Processing Systems*, volume 37, pp. 63082–63109, 2024. doi: 10.52202/079017-2014. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/7343a5c976f8399880b695267f1f9e9f-Abstract-Conference.html.

Christian Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete and Continuous Dynamical Systems*, 34(4):1533–1574, 2014. doi: 10.3934/dcds.2014.34.1533. URL https://doi.org/10.3934/dcds.2014.34.1533.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common objects in context. In *Computer Vision – ECCV 2014*, pp. 740–755, 2014. doi: 10.1007/978-3-319-10602-1_48.

Zhiqiu Lin, Deepak Pathak, Baiqi Li, Jiayao Li, Xide Xia, Graham Neubig, Pengchuan Zhang, and Deva Ramanan. Evaluating text-to-visual generation with image-to-text generation. In *Computer Vision – ECCV 2024*, pp. 366–384, 2025. doi: 10.1007/978-3-031-72673-6_20.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=XVjTT1nw5z.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL https://arxiv.org/abs/1711.05101.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver: A fast ODE solver for diffusion probabilistic model sampling in around 10 steps. In *Advances in Neural Information Processing Systems*, volume 35, pp. 5775–5787, 2022. doi: 10.52202/068431-0418.

Yunhong Min, Juil Koo, Seungwoo Yoo, and Minhyuk Sung. BézierFlow: Learning Bézier stochastic interpolant schedulers for few-step generation. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=PCuDI32xhQ.

Anton Obukhov, Maximilian Seitzer, Po-Wei Wu, Semen Zhydenko, Jonathan Kyl, and Elvis Yu-Jing Lin. High-fidelity performance metrics for generative models in PyTorch, 2020. URL https://github.com/toshas/torch-fidelity. Version 0.4.0; DOI: 10.5281/zenodo.3786539.

Xue Bin Peng, Aviral Kumar, Grace Zhang, and Sergey Levine. Advantage-weighted regression: Simple and scalable off-policy reinforcement learning. *arXiv preprint arXiv:1910.00177*, 2019. URL https://arxiv.org/abs/1910.00177.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Rombach_High-Resolution_Image_Synthesis_With_Latent_Diffusion_Models_CVPR_2022_paper.html.

Amirmojtaba Sabour, Sanja Fidler, and Karsten Kreis. Align your steps: Optimizing sampling schedules in diffusion models. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 42947–42975. PMLR, 2024. URL https://proceedings.mlr.press/v235/sabour24a.html.

Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training GANs. In *Advances in Neural Information Processing Systems*, volume 29, 2016. URL https://proceedings.neurips.cc/paper/2016/hash/8a3363abe792db2d8761d6403605aeb7-Abstract.html.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017. URL https://arxiv.org/abs/1707.06347.

Neta Shaul, Juan Perez, Ricky T. Q. Chen, Ali Thabet, Albert Pumarola, and Yaron Lipman. Bespoke solvers for generative flow models. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=1PXEY7ofFX.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021a. URL https://openreview.net/forum?id=St1giarCHLP.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021b. URL https://openreview.net/forum?id=PxTIG12RRHS.

Dejan Stancevic, Florian Handke, and Luca Ambrogioni. Entropic time schedulers for generative diffusion models. In *Advances in Neural Information Processing Systems*, volume 38, pp. 44222–44252, 2025. doi: 10.52202/085713-1474. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/3ed10616ecdd776be283c0a45cf9332d-Abstract-Conference.html.

Jianlin Su, Murtadha Ahmed, Yu Lu, Shengfeng Pan, Wen Bo, and Yunfeng Liu. RoFormer: Enhanced transformer with Rotary Position Embedding. *Neurocomputing*, 568:127063, 2024. doi: 10.1016/j.neucom.2023.127063.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=CD9Snc73AW.

Vinh Tong, Trung-Dung Hoang, Anji Liu, Guy Van den Broeck, and Mathias Niepert. Learning to discretize denoising diffusion ODEs. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/750046157471c56235a781f2eff6e226-Abstract-Conference.html.

Brandon Trabucco, Aviral Kumar, Xinyang Geng, and Sergey Levine. Conservative objective models for effective offline model-based optimization. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 10358–10368. PMLR, 2021. URL https://proceedings.mlr.press/v139/trabucco21a.html.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

Shuai Wang, Zexian Li, Qipeng Zhang, Tianhui Song, Xubin Li, Tiezheng Ge, Bo Zheng, and Limin Wang. Differentiable solver search for fast diffusion sampling. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 63549–63568. PMLR, 2025. URL https://proceedings.mlr.press/v267/wang25bh.html.

Christopher Williams, Andrew Campbell, Arnaud Doucet, and Saifuddin Syed. Score-optimal diffusion schedules. In *Advances in Neural Information Processing Systems*, volume 37, pp. 107960–107983, 2024. doi: 10.52202/079017-3428. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/c34057f0cfb685eadf7d8d7cc99f734a-Abstract-Conference.html.

Enze Xie, Junsong Chen, Junyu Chen, Han Cai, Haotian Tang, Yujun Lin, Zhekai Zhang, Muyang Li, Ligeng Zhu, Yao Lu, and Song Han. SANA: Efficient high-resolution text-to-image synthesis with linear diffusion transformers. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/34345e243156da67605d4b63d71c8d98-Abstract-Conference.html.

Jiazheng Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, and Yuxiao Dong. ImageReward: Learning and evaluating human preferences for text-to-image generation. In *Advances in Neural Information Processing Systems*, volume 36, pp. 15903–15935, 2023. doi: 10.52202/075280-0700.

Shuchen Xue, Zhaoqiang Liu, Fei Chen, Shifeng Zhang, Tianyang Hu, Enze Xie, and Zhenguo Li. Accelerating diffusion sampling with optimized time steps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8292–8301, 2024. doi: 10.1109/CVPR52733.2024.00792.

Zilyu Ye, Zhiyang Chen, Tiancheng Li, Zemin Huang, Weijian Luo, and Guo-Jun Qi. Schedule on the fly: Diffusion time prediction for faster and better image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 23412–23422, 2025. URL https://openaccess.thecvf.com/content/CVPR2025/html/Ye_Schedule_On_the_Fly_Diffusion_Time_Prediction_for_Faster_and_CVPR_2025_paper.html.

Peiyu Yu, Suraj Kothawade, Sirui Xie, Ying Nian Wu, and Hongliang Fei. Designing instance-level sampling schedules via REINFORCE with James–Stein shrinkage. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 36040–36050, 2026. URL https://openaccess.thecvf.com/content/CVPR2026/html/Yu_Designing_Instance-Level_Sampling_Schedules_via_REINFORCE_with_James-Stein_Shrinkage_CVPR_2026_paper.html.

Qinsheng Zhang and Yongxin Chen. Fast sampling of diffusion models with exponential integrator. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=Loek7hfb46P.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018. URL https://openaccess.thecvf.com/content_cvpr_2018/html/Zhang_The_Unreasonable_Effectiveness_CVPR_2018_paper.html.

Wenliang Zhao, Lujia Bai, Yongming Rao, Jie Zhou, and Jiwen Lu. UniPC: A unified predictor-corrector framework for fast sampling of diffusion models. In *Advances in Neural Information Processing Systems*, volume 36, pp. 49842–49869, 2023. doi: 10.52202/075280-2170.

A SUPPORTING PROOFS AND DERIVATIONS

This appendix supports the representation and utility tilt in Sec. 3, the learning targets and refinement in Sec. 4, and the allocation analysis in Sec. 5. It gives exact finite-budget defect identities and conditional improvement bounds, their asymptotic allocation specialization, utility-tilt and Doob characterizations, and the finite-target and stochastic-policy details. The population clock laws, empirical allocation anchors, and refined policies are distinguished throughout.

A.1 EXACT RE-CLOCKING AND FIXED-FIELD SCOPE

For clock time λ and $\boldsymbol{y}_\lambda = \boldsymbol{x}_{\phi_\rho(\lambda)}$, the transformed ODE is

$$\frac{\mathrm{d}\boldsymbol{y}_\lambda}{\mathrm{d}\lambda} = \frac{\boldsymbol{v}_\theta(\boldsymbol{y}_\lambda, \phi_\rho(\lambda); \boldsymbol{c})}{\rho(\phi_\rho(\lambda))}. \tag{16}$$

Lemma 1 (Exact re-clocking)**.** *Let $\rho \in \mathcal{R}$ satisfy $0 < \rho_{\min} \le \rho(t) \le \rho_{\max} < \infty$ almost everywhere. Define $F_\rho(t) = \int_0^t \rho(a)\,\mathrm{d}a$ and $\phi_\rho = F_\rho^{-1}$. If $\boldsymbol{x}_t$ solves Eq. (1), then $\boldsymbol{y}_\lambda = \boldsymbol{x}_{\phi_\rho(\lambda)}$ is absolutely continuous and satisfies Eq. (16) almost everywhere. In particular, $\boldsymbol{y}_1 = \boldsymbol{x}_1$.*

Proof. The bounds on ρ make F_ρ strictly increasing, absolutely continuous, and bi-Lipschitz from $[0, 1]$ onto itself. Its inverse is absolutely continuous and obeys

$$\phi'_\rho(\lambda) = \frac{1}{\rho(\phi_\rho(\lambda))} \quad \text{for a.e. } \lambda.$$

The chain rule for absolutely continuous functions gives

$$\frac{\mathrm{d}\boldsymbol{y}_\lambda}{\mathrm{d}\lambda} = \boldsymbol{v}_\theta(\boldsymbol{x}_{\phi_\rho(\lambda)}, \phi_\rho(\lambda); \boldsymbol{c})\phi'_\rho(\lambda) = \frac{\boldsymbol{v}_\theta(\boldsymbol{y}_\lambda, \phi_\rho(\lambda); \boldsymbol{c})}{\rho(\phi_\rho(\lambda))}$$

almost everywhere. Since $F_\rho(1) = 1$, $\phi_\rho(1) = 1$, so the endpoint is unchanged. □

This result includes the positive piecewise-constant densities used in the finite-bin implementation. Reparameterization preserves the exact endpoint; the implementation evaluates the original frozen field on the induced grid, as specified in Sec. 3.

A.2 FINITE-STEP NON-INVARIANCE: AN EULER WITNESS

For a smooth frozen field, the exact-minus-Euler local defect on an induced interval is

$$\boldsymbol{\delta}^\rho_{\mathrm{FE},k} = \frac{(\Delta t^\rho_k)^2}{2}(\mathrm{D}_t\boldsymbol{v}_\theta)(\boldsymbol{x}_{\tau^\rho_k}, \tau^\rho_k; \boldsymbol{c}) + O((\Delta t^\rho_k)^3). \tag{17}$$

Here $\mathrm{D}_t = \partial_t + \boldsymbol{v}_\theta \cdot \nabla_{\boldsymbol{x}}$, with $\mathrm{D}_t\boldsymbol{v}_\theta = \partial_t\boldsymbol{v}_\theta + (\boldsymbol{J}_{\boldsymbol{x}}\boldsymbol{v}_\theta)\boldsymbol{v}_\theta$. We use column vectors and $[\boldsymbol{J}_{\boldsymbol{x}}\boldsymbol{f}]_{ij} = \partial f_i/\partial x_j$. The interval width and its location depend on the clock. We use the Jacobian convention of Sec. 3.

Lemma 2 (Forward-Euler local defect)**.** *Assume $\boldsymbol{v}_\theta$ is twice continuously differentiable on a neighborhood of the exact trajectory. For a step of width $\Delta t > 0$ beginning at $(\boldsymbol{x}_t, t)$, the exact-minus-Euler local defect is*

$$\Phi_{t,t+\Delta t}(\boldsymbol{x}_t) - \{\boldsymbol{x}_t + \Delta t\, \boldsymbol{v}_\theta(\boldsymbol{x}_t, t; \boldsymbol{c})\} = \frac{(\Delta t)^2}{2}(\mathrm{D}_t\boldsymbol{v}_\theta)(\boldsymbol{x}_t, t; \boldsymbol{c}) + O((\Delta t)^3),$$

where

$$\mathrm{D}_t := \partial_t + \boldsymbol{v}_\theta(\boldsymbol{x}, t; \boldsymbol{c})\cdot\nabla_{\boldsymbol{x}}, \qquad (\mathrm{D}_t\boldsymbol{v}_\theta)(\boldsymbol{x}, t; \boldsymbol{c}) = \partial_t\boldsymbol{v}_\theta(\boldsymbol{x}, t; \boldsymbol{c}) + \boldsymbol{J}_{\boldsymbol{x}}\boldsymbol{v}_\theta(\boldsymbol{x}, t; \boldsymbol{c})\, \boldsymbol{v}_\theta(\boldsymbol{x}, t; \boldsymbol{c}).$$

The operator D_t is the material derivative induced by the frozen field.

Proof. Along the solution, $\dot{\boldsymbol{x}}_t = \boldsymbol{v}_\theta(\boldsymbol{x}_t, t; \boldsymbol{c})$ and

$$\ddot{\boldsymbol{x}}_t = \partial_t\boldsymbol{v}_\theta(\boldsymbol{x}_t, t; \boldsymbol{c}) + \boldsymbol{J}_{\boldsymbol{x}}\boldsymbol{v}_\theta(\boldsymbol{x}_t, t; \boldsymbol{c})\dot{\boldsymbol{x}}_t = (\mathrm{D}_t\boldsymbol{v}_\theta)(\boldsymbol{x}_t, t; \boldsymbol{c}).$$

Taylor expansion of $\boldsymbol{x}_{t+\Delta t}$ about t gives the result. □

For $\Delta t = \Delta t^\rho_k$ and $t = \tau^\rho_k$, this is the defect in Eq. (17); both its width and evaluation point depend on ρ. The resulting endpoint dependence is generic rather than universal: the coefficient can vanish, a method can be exact for a special field, or propagated signed defects can cancel.

A.3 Node density induced by inverse-CDF placement

Lemma 3 (Clock-induced step widths)**.** *Let $0 < \rho_{\min} \le \rho(t) \le \rho_{\max} < \infty$ almost everywhere and set $\tau_k^\rho = F_\rho^{-1}(k/K)$. Then*

$$\frac{1}{K\rho_{\max}} \le \Delta t_k^\rho := \tau_{k+1}^\rho - \tau_k^\rho \le \frac{1}{K\rho_{\min}}.$$

If ρ is continuous on $[\tau_k^\rho, \tau_{k+1}^\rho]$, there is a ζ_k^ρ in that interval such that

$$\Delta t_k^\rho = \frac{1}{K\rho(\zeta_k^\rho)}.$$

Moreover, for every continuous f,

$$\frac{1}{K}\sum_{k=0}^{K-1} f(\tau_k^\rho) \longrightarrow \int_0^1 f(t)\rho(t)\,\mathrm{d}t.$$

Proof. Inverse-CDF placement gives

$$\int_{\tau_k^\rho}^{\tau_{k+1}^\rho} \rho(t)\,\mathrm{d}t = \frac{1}{K}.$$

The density bounds imply the two-sided step-width bound, and continuity on the interval gives the stated ζ_k^ρ by the integral mean-value theorem. For the final claim, with $\lambda_k = k/K$,

$$\frac{1}{K}\sum_{k=0}^{K-1} f(\tau_k^\rho) = \frac{1}{K}\sum_{k=0}^{K-1} f(\phi_\rho(\lambda_k)) \longrightarrow \int_0^1 f(\phi_\rho(\lambda))\,\mathrm{d}\lambda = \int_0^1 f(t)\rho(t)\,\mathrm{d}t.$$

□

Thus $K\rho(t)\,\mathrm{d}t$ is the asymptotic number of macro-steps assigned near t, and the uniform lower density bound gives $\max_k \Delta t_k^\rho = O(K^{-1})$. For multistage solvers, the τ_k^ρ are macro-step boundaries; internal stage times remain those prescribed by the fixed solver on each interval.

A.4 Finite-budget defects and asymptotic allocation analysis

Let $\Phi_{t_0,t_1}(\boldsymbol{x})$ denote the exact flow map of Eq. (1) with fixed context suppressed, and let $\Psi_{s,\Delta t}(\boldsymbol{x}, t; \boldsymbol{c})$ denote one macro-step of solver s. On the clock grid, the full exact-minus-numerical local defect is

$$\boldsymbol{\delta}_{s,k}^\rho = \Phi_{\tau_k^\rho,\tau_{k+1}^\rho}(\boldsymbol{x}_{\tau_k^\rho}) - \Psi_{s,\Delta t_k^\rho}(\boldsymbol{x}_{\tau_k^\rho}, \tau_k^\rho; \boldsymbol{c}). \tag{18}$$

Exact finite-budget propagation. Fix a finite grid with $K = K_s(b) \ge 1$, and abbreviate $\boldsymbol{x}_k = \boldsymbol{x}_{\tau_k^\rho}$, $\boldsymbol{y}_k = \widehat{\boldsymbol{x}}_{\tau_k^\rho}^\rho$, $\Psi_k(\boldsymbol{x}) = \Psi_{s,\Delta t_k^\rho}(\boldsymbol{x}, \tau_k^\rho; \boldsymbol{c})$, and $\boldsymbol{\delta}_k = \boldsymbol{\delta}_{s,k}^\rho$. Thus $\boldsymbol{y}_{k+1} = \Psi_k(\boldsymbol{y}_k)$ and $\boldsymbol{y}_0 = \boldsymbol{x}_0$; $\boldsymbol{y}_K$ and $\boldsymbol{x}_K$ are the numerical and exact terminal states. Write $u(\boldsymbol{x}) = u_m(\boldsymbol{x}, \boldsymbol{c})$. The following integral mean-value argument specializes classical adjoint-based error analysis (Cao & Petzold, 2004) to the full clock-dependent numerical updates; it does not truncate the local defect.

Proposition 1 (Exact finite-budget utility decomposition)**.** *Assume each Ψ_k is continuously differentiable on a neighborhood of the segment joining $\boldsymbol{x}_k$ and $\boldsymbol{y}_k$, and u is continuously differentiable on a neighborhood of the terminal segment. Set $\boldsymbol{e}_k = \boldsymbol{y}_k - \boldsymbol{x}_k$ and define*

$$\begin{aligned} \overline{\boldsymbol{A}}_k &= \int_0^1 \boldsymbol{J}_{\boldsymbol{x}}\Psi_k(\boldsymbol{x}_k + \theta\boldsymbol{e}_k)\,\mathrm{d}\theta, \\ \overline{\boldsymbol{a}}_K &= \int_0^1 \nabla u(\boldsymbol{x}_K + \theta\boldsymbol{e}_K)\,\mathrm{d}\theta, \qquad \overline{\boldsymbol{a}}_k = \overline{\boldsymbol{A}}_k^\top \overline{\boldsymbol{a}}_{k+1}. \end{aligned} \tag{19}$$

Then, for this finite grid,

$$u(\boldsymbol{y}_K) - u(\boldsymbol{x}_K) = -\sum_{k=0}^{K-1} \langle \overline{\boldsymbol{a}}_{k+1}, \boldsymbol{\delta}_k \rangle. \tag{20}$$

No small-step or solver-order assumption is required.

Proof. The integral mean-value formula gives $\boldsymbol{e}_{k+1} = \overline{\boldsymbol{A}}_k \boldsymbol{e}_k - \boldsymbol{\delta}_k$ and $u(\boldsymbol{y}_K) - u(\boldsymbol{x}_K) = \overline{\boldsymbol{a}}_K^\top \boldsymbol{e}_K$. Hence

$$\overline{\boldsymbol{a}}_{k+1}^\top \boldsymbol{e}_{k+1} = \overline{\boldsymbol{a}}_k^\top \boldsymbol{e}_k - \overline{\boldsymbol{a}}_{k+1}^\top \boldsymbol{\delta}_k.$$

Summing and using $\boldsymbol{e}_0 = 0$ proves the identity. □

Both the full defects and the averaged discrete adjoints depend on the finite clock and its numerical trajectory. Their dependence retains nonlinear propagation and large-step effects; Eq. (20) does not identify a clock-independent profile or make these quantities available from offline utility observations. A bound for an ordinary discrete adjoint makes the effect of linearization explicit. All norms below are Euclidean or the corresponding induced operator and bilinear norms.

Proposition 2 (Finite-budget remainder and allocation comparison)**.** *Under the preceding setup, suppose the maps and utility are C^2 on the stated neighborhoods. On the respective state segments, let L_k and H_k bound $\|\boldsymbol{J}_{\boldsymbol{x}} \Psi_k\|$ and $\|\nabla^2_{\boldsymbol{x}} \Psi_k\|$, and let H_u bound $\|\nabla^2 u\|$ on the terminal segment. Define*

$$\begin{aligned} \boldsymbol{a}_K^{\mathrm{d}} &= \nabla u(\boldsymbol{x}_K), & \boldsymbol{a}_k^{\mathrm{d}} &= \boldsymbol{J}_{\boldsymbol{x}} \Psi_k(\boldsymbol{x}_k)^\top \boldsymbol{a}_{k+1}^{\mathrm{d}}, \\ E_0 &= 0, & E_{k+1} &= L_k E_k + \|\boldsymbol{\delta}_k\|, \\ \mathcal{D}_K(\rho) &= \sum_{k=0}^{K-1} \langle \boldsymbol{a}_{k+1}^{\mathrm{d}}, \boldsymbol{\delta}_k \rangle, & \mathcal{B}_K(\rho) &= \frac{H_u}{2} E_K^2 + \frac{1}{2} \sum_{k=0}^{K-1} \|\boldsymbol{a}_{k+1}^{\mathrm{d}}\| H_k E_k^2. \end{aligned} \tag{21}$$

Then

$$\left| u(\boldsymbol{y}_K) - u(\boldsymbol{x}_K) + \mathcal{D}_K(\rho) \right| \le \mathcal{B}_K(\rho). \tag{22}$$

For two clocks ρ, ρ_0 at the same setting and source draw, it follows that

$$u_m(\widehat{\boldsymbol{x}}_1^{\rho}, \boldsymbol{c}) - u_m(\widehat{\boldsymbol{x}}_1^{\rho_0}, \boldsymbol{c}) \ge \mathcal{D}_K(\rho_0) - \mathcal{D}_K(\rho) - \mathcal{B}_K(\rho_0) - \mathcal{B}_K(\rho). \tag{23}$$

A strictly positive right-hand side is sufficient for finite-budget utility improvement. If all update maps are affine and u is linear, we can take $H_k = H_u = 0$, so $\mathcal{B}_K = 0$ and Eq. (22) is an equality with zero residual at every budget.

Proof. The Jacobian bounds give $\|\boldsymbol{e}_{k+1}\| \le L_k \|\boldsymbol{e}_k\| + \|\boldsymbol{\delta}_k\|$, hence $\|\boldsymbol{e}_k\| \le E_k$. Taylor's formula with its remainder retained gives

$$\boldsymbol{e}_{k+1} = \boldsymbol{J}_{\boldsymbol{x}} \Psi_k(\boldsymbol{x}_k) \boldsymbol{e}_k - \boldsymbol{\delta}_k + \boldsymbol{r}_k, \qquad \|\boldsymbol{r}_k\| \le \tfrac{1}{2} H_k E_k^2.$$

Telescoping with $\boldsymbol{a}_k^{\mathrm{d}}$, and writing the terminal Taylor remainder as r_u, yields

$$u(\boldsymbol{y}_K) - u(\boldsymbol{x}_K) = -\mathcal{D}_K(\rho) + \sum_{k=0}^{K-1} \langle \boldsymbol{a}_{k+1}^{\mathrm{d}}, \boldsymbol{r}_k \rangle + r_u, \qquad |r_u| \le \tfrac{1}{2} H_u E_K^2.$$

The triangle inequality proves Eq. (22). The exact endpoint is common to both clocks, so subtraction proves Eq. (23). Affine updates and linear utility have zero Hessians. □

These are conditional analytical bounds, not computed certificates for the pretrained-model experiments. Large defects, Jacobian products, or curvature can make $\mathcal{B}_K$ uninformative, even though the identity remains exact. A differentiable batch utility is covered by the product state; conditional expectations preserve the identities and inequalities when their terms are integrable. Nonsmooth scores need not admit this adjoint representation. For a fixed multistep rule, the finite-budget argument applies to an augmented history state when its update maps satisfy the stated differentiability conditions and the full startup defect is included. This does not by itself establish the classical-order expansion for variable-step AB2: its Euler startup has an $O((\Delta t_0)^2)$ defect, of the same order as an $O(K^{-2})$ global error, and must be retained in any second-order expansion.

Small-mesh specialization. The next theorem replaces the full defects by their principal terms and the discrete propagation by the exact-flow Jacobian. These additional approximations, not Eq. (20), require mesh refinement.

Theorem 1 (Asymptotic adjoint-weighted terminal defect)**.** *Fix a source draw $\boldsymbol{x}_0$ and $(m, s, \boldsymbol{c})$, and let $K \to \infty$ through budgets with $K_s(b_K) = K$. Assume all clocks in the sequence satisfy common bounds $0 < \rho_{\min} \le \rho(t) \le \rho_{\max} < \infty$. Assume the frozen field and a scalar samplewise utility $u_m(\cdot, \boldsymbol{c})$ are sufficiently smooth, the one-step map is stable and C^2 in state with $\|\nabla^2_{\boldsymbol{x}} \Psi_{s,\Delta t}\| = O(\Delta t)$ near the trajectory, and the following local-defect expansion holds uniformly together with its state Jacobian:*

$$\boldsymbol{\delta}^{\rho}_{s,k} = \boldsymbol{d}_s(\boldsymbol{x}_{\tau^{\rho}_k}, \tau^{\rho}_k; \boldsymbol{c})(\Delta t^{\rho}_k)^{p_s+1} + O((\Delta t^{\rho}_k)^{p_s+2}), \tag{24}$$

where $p_s \in \mathbb{N}$ is the classical order. If $\widehat{\boldsymbol{x}}^{\rho}_1$ is the numerical endpoint, then

$$\widehat{\boldsymbol{x}}^{\rho}_1 - \boldsymbol{x}_1 = -\sum_{k=0}^{K-1} \boldsymbol{J}_{\boldsymbol{x}} \Phi_{\tau^{\rho}_{k+1},1}(\boldsymbol{x}_{\tau^{\rho}_{k+1}}) \boldsymbol{d}_s(\boldsymbol{x}_{\tau^{\rho}_k}, \tau^{\rho}_k; \boldsymbol{c})(\Delta t^{\rho}_k)^{p_s+1} + o(K^{-p_s}), \tag{25}$$

and Eq. (13) follows with

$$\boldsymbol{a}_m(t) = \left[\boldsymbol{J}_{\boldsymbol{x}} \Phi_{t,1}(\boldsymbol{x}_t)\right]^{\top} \nabla_{\boldsymbol{x}} u_m(\boldsymbol{x}_1, \boldsymbol{c}).$$

Equivalently,

$$-\dot{\boldsymbol{a}}_m(t) = \left[\boldsymbol{J}_{\boldsymbol{x}} \boldsymbol{v}_\theta(\boldsymbol{x}_t, t; \boldsymbol{c})\right]^{\top} \boldsymbol{a}_m(t), \qquad \boldsymbol{a}_m(1) = \nabla_{\boldsymbol{x}} u_m(\boldsymbol{x}_1, \boldsymbol{c}).$$

Proof. Let $\widehat{\boldsymbol{x}}^{\rho}_{\tau^{\rho}_k}$ be the numerical state and $\boldsymbol{e}^{\rho}_k = \widehat{\boldsymbol{x}}^{\rho}_{\tau^{\rho}_k} - \boldsymbol{x}_{\tau^{\rho}_k}$. Differentiating the uniformly valid defect expansion in the state gives

$$\boldsymbol{J}_{\boldsymbol{x}} \Psi_{s,\Delta t^{\rho}_k}(\boldsymbol{x}_{\tau^{\rho}_k}, \tau^{\rho}_k; \boldsymbol{c}) = \boldsymbol{J}_{\boldsymbol{x}} \Phi_{\tau^{\rho}_k, \tau^{\rho}_{k+1}}(\boldsymbol{x}_{\tau^{\rho}_k}) + O((\Delta t^{\rho}_k)^{p_s+1}).$$

Taylor expansion of the exact error recursion in Proposition 1 therefore yields

$$\boldsymbol{e}^{\rho}_{k+1} = \boldsymbol{J}_{\boldsymbol{x}} \Phi_{\tau^{\rho}_k, \tau^{\rho}_{k+1}}(\boldsymbol{x}_{\tau^{\rho}_k}) \boldsymbol{e}^{\rho}_k - \boldsymbol{\delta}^{\rho}_{s,k} + \boldsymbol{R}^{\rho}_{s,k}.$$

Stability and Lemma 3 give $\max_k \Delta t^{\rho}_k = O(K^{-1})$ and $\max_k \|\boldsymbol{e}^{\rho}_k\| = O(K^{-p_s})$. Together with the Jacobian-level defect remainder and the second-derivative bound, this yields

$$\sum_{k=0}^{K-1} O\big(\Delta t^{\rho}_k \|\boldsymbol{e}^{\rho}_k\|^2 + (\Delta t^{\rho}_k)^{p_s+1} \|\boldsymbol{e}^{\rho}_k\| + (\Delta t^{\rho}_k)^{p_s+2}\big) = o(K^{-p_s}).$$

Discrete variation of constants then gives Eq. (25); the minus sign follows from defining the defect as exact minus numerical. Finally,

$$u_m(\widehat{\boldsymbol{x}}^{\rho}_1, \boldsymbol{c}) - u_m(\boldsymbol{x}_1, \boldsymbol{c}) = \nabla_{\boldsymbol{x}} u_m(\boldsymbol{x}_1, \boldsymbol{c})^{\top} (\widehat{\boldsymbol{x}}^{\rho}_1 - \boldsymbol{x}_1) + o(K^{-p_s}),$$

which yields Eq. (13). Differentiating the flow-map Jacobian representation gives the adjoint equation. □

Replacing $\boldsymbol{a}_m(\tau^{\rho}_{k+1})$ by $\boldsymbol{a}_m(\tau^{\rho}_k)$ changes the sum by $O(K^{-(p_s+1)}) = o(K^{-p_s})$: the adjoint is Lipschitz and $\max_k \Delta t^{\rho}_k = O(K^{-1})$. The expansion is samplewise and signed; a differentiable batch utility follows by applying the argument to the product ODE. Eq. (14) replaces this signed expansion by a nonnegative allocation surrogate and therefore discards cancellation. For smooth utilities, $\kappa_\xi(t)$ may be interpreted as an average or envelope over source and evaluation draws of

$$\left|\langle \boldsymbol{a}_m(t), \boldsymbol{d}_s(\boldsymbol{x}_t, t; \boldsymbol{c})\rangle\right|.$$

The exact finite-budget terms in Eq. (20) generally depend on the clock, so they need not reduce to this fixed profile. For nonsmooth scores and very low NFEs, a clock-independent κ_ξ, including a budget-dependent one, remains a modeling choice rather than a consequence of the identity. GICO instead fits the measured finite-budget utility.

A.5 ALLOCATION SURROGATE AND OPTIMAL DENSITY

Theorem 2 (Allocation surrogate and optimal density)**.** *Let* $p_s \in \mathbb{N}$.

(i) Continuum limit. *If* ρ *and* κ_ξ *are continuous,* $0 < \rho_{\min} \le \rho \le \rho_{\max} < \infty$*, and the profile is fixed along the refinement sequence, then*

$$\mathcal{A}^{\mathrm{disc}}_{\xi,K}(\rho) = \sum_{k=0}^{K-1} \kappa_\xi(\tau_k^\rho)(\Delta t_k^\rho)^{p_s+1}$$

satisfies

$$K^{p_s}\mathcal{A}^{\mathrm{disc}}_{\xi,K}(\rho) \longrightarrow \int_0^1 \kappa_\xi(t)\rho(t)^{-p_s}\,\mathrm{d}t. \tag{26}$$

(ii) Optimal allocation. *If* $\kappa_\xi(t) > 0$ *almost everywhere and* $\kappa_\xi^{1/(p_s+1)} \in L^1([0,1])$*, then Eq. (15) is the unique minimizer over* $\mathcal{R}$*, and*

$$\min_{\rho\in\mathcal{R}} \int_0^1 \kappa_\xi(t)\rho(t)^{-p_s}\,\mathrm{d}t = \left(\int_0^1 \kappa_\xi(t)^{1/(p_s+1)}\,\mathrm{d}t\right)^{p_s+1}. \tag{27}$$

Proof. For part (i), let $\lambda_k = k/K$ and $\phi = \phi_\rho$. For some $\bar\lambda_k \in (\lambda_k, \lambda_{k+1})$,

$$K\Delta t_k^\rho = K\{\phi(\lambda_{k+1}) - \phi(\lambda_k)\} = \phi'(\bar\lambda_k).$$

Therefore

$$K^{p_s}\mathcal{A}^{\mathrm{disc}}_{\xi,K}(\rho) = \frac{1}{K}\sum_{k=0}^{K-1} \kappa_\xi(\phi(\lambda_k))\phi'(\bar\lambda_k)^{p_s+1},$$

which converges to

$$\int_0^1 \kappa_\xi(\phi(\lambda))\phi'(\lambda)^{p_s+1}\,\mathrm{d}\lambda = \int_0^1 \kappa_\xi(t)\rho(t)^{-p_s}\,\mathrm{d}t.$$

The same limit holds for bounded positive piecewise-continuous densities with finitely many jumps: apply the argument on each continuity interval; after rescaling, the finitely many grid cells crossing jumps contribute $O(K^{-1})$. For part (ii), Hölder's inequality gives

$$\begin{aligned}\int_0^1 \kappa_\xi(t)^{1/(p_s+1)}\,\mathrm{d}t &= \int_0^1 \{\kappa_\xi(t)\rho(t)^{-p_s}\}^{1/(p_s+1)}\rho(t)^{p_s/(p_s+1)}\,\mathrm{d}t \\ &\le \left(\int_0^1 \kappa_\xi(t)\rho(t)^{-p_s}\,\mathrm{d}t\right)^{1/(p_s+1)}\left(\int_0^1 \rho(t)\,\mathrm{d}t\right)^{p_s/(p_s+1)}.\end{aligned}$$

Since $\int \rho = 1$, this proves the lower bound in Eq. (27). Equality holds exactly when $\rho(t)^{p_s+1} \propto \kappa_\xi(t)$, and normalization gives Eq. (15). Strict convexity of $u \mapsto u^{-p_s}$, together with $\kappa_\xi > 0$ almost everywhere, gives uniqueness. □

The Fermat-type analogy concerns allocation along a fixed trajectory: Eq. (15) makes each interval's integral of $\kappa_\xi(t)^{1/(p_s+1)}$ equal to $K^{-1}\int_0^1 \kappa_\xi(t)^{1/(p_s+1)}\,\mathrm{d}t$. This is a weighted-coordinate interpretation, not optical path optimization. For a profile fixed along refinement, K^{-p_s} changes the surrogate value but not its minimizer. At low NFEs, budget-dependent profiles are a modeling choice; the analysis does not guarantee budget transfer. If κ_ξ vanishes, the unconstrained optimizer may lie on the boundary of $\mathcal{R}$. For integrable $\kappa_\xi > 0$ almost everywhere and analytic box constraints $0 < \rho_{\min} < 1 < \rho_{\max} < \infty$, the convex KKT solution is

$$\rho^\star_{\xi,\mathrm{box}}(t) = \mathrm{clip}\left(\gamma_{\xi,\mathrm{box}}\kappa_\xi(t)^{1/(p_s+1)}, \rho_{\min}, \rho_{\max}\right),$$

where $\gamma_{\xi,\mathrm{box}} > 0$ normalizes the density. If either bound equals one, uniform is the only feasible density. If $\kappa_\xi = 0$ on a set of positive measure, the clipped formula fixes $\rho = \rho_{\min}$ there. If it cannot then be normalized, the remaining mass is allocated on that zero-cost set within the box, and

minimizers need not be unique there. The finite-bin numerical guard in Appx B.1 separately enforces positive, finite representations; it does not require or estimate κ_ξ.

Controlled validation of the utility-dependent allocation. We specialize Propositions 1–2 and Theorems 1–2 to an exactly solvable two-dimensional ODE. The dynamics, source draws, and Euler update rule are held fixed while the terminal utility changes. These checks concern analytically constructed allocations. For $i \in \{1, 2\}$, let

$$\dot{x}_{t,i} = g_i(t)x_{t,i}, \qquad g_i(t) = 0.05 + 0.2\,\sigma\big(20(t - \zeta_i)\big), \qquad (\zeta_1, \zeta_2) = (0.25, 0.75),$$

where $\sigma(z) = (1 + e^{-z})^{-1}$. Initial states follow $\boldsymbol{x}_0 \sim \mathrm{Uniform}([0.8, 1.2]^2)$, with zero conditioning, and $u_1(\boldsymbol{x}) = x_1$, $u_2(\boldsymbol{x}) = x_2$. The first coefficient changes most sharply early and the second late.

Exact trajectory, adjoint, and allocation. Define the integrated coefficient

$$G_i(t) = \int_0^t g_i(a)\,\mathrm{d}a = 0.05t + 0.01 \log\left(\frac{1 + e^{20(t-\zeta_i)}}{1 + e^{-20\zeta_i}}\right).$$

Then $x_{t,i} = x_{0,i}e^{G_i(t)}$. For utility u_m, the terminal condition is the coordinate vector $\boldsymbol{e}_m$, and the backward adjoint has the exact solution

$$\boldsymbol{a}_m(t) = e^{G_m(1)-G_m(t)}\boldsymbol{e}_m, \qquad -\dot{\boldsymbol{a}}_m(t) = \mathrm{diag}(g_1(t), g_2(t))\boldsymbol{a}_m(t), \qquad \boldsymbol{a}_m(1) = \boldsymbol{e}_m.$$

Writing $q_i(t) = g_i'(t) + g_i(t)^2$, with $g_i'(t) = 4\sigma(20(t - \zeta_i))[1 - \sigma(20(t - \zeta_i))]$, the principal exact-minus-Euler defect is $d_{\mathrm{FE},i}(\boldsymbol{x}_t, t) = \frac{1}{2}q_i(t)x_{t,i}$. For a fixed source draw, its adjoint-weighted profile therefore simplifies to

$$\kappa_m(t) := \langle \boldsymbol{a}_m(t), \boldsymbol{d}_{\mathrm{FE}}(\boldsymbol{x}_t, t)\rangle = \frac{x_{1,m}}{2}q_m(t) > 0, \qquad \rho_m^\star(t) = \frac{\sqrt{q_m(t)}}{\int_0^1 \sqrt{q_m(a)}\,\mathrm{d}a}.$$

The positive factor $x_{1,m}/2$ cancels in normalized density. Thus the allocation is source-independent but utility-dependent: the two profiles are not proportional, so their unique surrogate minimizers differ. A fixed utility-independent defect-norm allocation cannot minimize both allocation surrogates. Positivity also makes all leading contributions sign-consistent; no absolute-value relaxation is needed here.

Finite-bin construction and surrogate check. Use $B = 64$ equal flow-time bins $I_j = [j/B, (j + 1)/B)$, with the final bin including one. Bin-integrating the continuous adjoint allocation gives

$$p_{m,j} = \int_{I_j} \rho_m^\star(t)\,\mathrm{d}t, \qquad \rho_{m,B}(t) = Bp_{m,j} \quad (t \in I_j), \qquad \tau_k^\rho = F_\rho^{-1}(k/K).$$

All masses are strictly positive. The represented density is fixed across budgets; only its decoded node count changes. With $p_s = 1$, the surrogate and its ratio to the continuous minimum are

$$\mathcal{A}_{m,K}(\rho) = \frac{x_{1,m}}{2K}\int_0^1 \frac{q_m(t)}{\rho(t)}\,\mathrm{d}t, \qquad \mathrm{AR}_m(\rho) = \frac{\displaystyle\int_0^1 q_m(t)/\rho(t)\,\mathrm{d}t}{\displaystyle\left(\int_0^1 \sqrt{q_m(t)}\,\mathrm{d}t\right)^2} \geq 1.$$

For $\rho_{m,B}$, the numerator is $\sum_j (Bp_{m,j})^{-1}\int_{I_j} q_m(t)\,\mathrm{d}t$. The ratios are 1.0004844 for u_1 and 1.0006703 for u_2, both within 0.07% of the continuous minimum. This assesses the finite representation of the analytical allocation, not finite-NFE optimality.

Finite-budget Euler utility. Forward Euler uses one field evaluation per interval, so $b = K$. For either $\rho \equiv 1$ or $\rho = \rho_{m,B}$, put $h_k = \tau_{k+1}^\rho - \tau_k^\rho$. Its endpoint and positive exact-minus-numerical utility error are

$$\widehat{x}_{1,m}^\rho = x_{0,m}\prod_{k=0}^{K-1}\big[1 + h_k g_m(\tau_k^\rho)\big], \qquad L_{m,K}(\rho) = x_{1,m} - \widehat{x}_{1,m}^\rho > 0.$$

Exact finite-budget allocation. The finite-step loss has the equivalent logarithmic form

$$\Lambda_{m,K}(\rho) = G_m(1) - \sum_{k=0}^{K-1}\log\big(1 + h_k g_m(\tau_k^\rho)\big), \qquad L_{m,K}(\rho) = x_{1,m}\big(1 - e^{-\Lambda_{m,K}(\rho)}\big). \tag{28}$$

Because $g_m > 0$ and $g'_m > 0$, each exact integrated coefficient exceeds $h_k g_m(\tau_k^\rho) > \log(1 + h_k g_m(\tau_k^\rho))$. Thus $\Lambda_{m,K} > 0$, and decreasing it is exactly equivalent to decreasing terminal-utility error and increasing u_m, at any finite K. For $f_{m,k} = 1 + h_k g_m(\tau_k^\rho)$, the full defect and discrete adjoint are

$$\delta_{m,k} = x_{\tau_k^\rho,m}\left[e^{G_m(\tau_{k+1}^\rho)-G_m(\tau_k^\rho)} - f_{m,k}\right], \qquad a^{\mathrm{d}}_{m,k+1} = \prod_{\ell=k+1}^{K-1} f_{m,\ell}.$$

An empty product is one. Euler's update is linear in state and both utilities are linear, so Proposition 2 gives the exact identity $L_{m,K} = \sum_k a^{\mathrm{d}}_{m,k+1}\delta_{m,k}$ with no linearization remainder. It is distinct from the principal-defect prediction below.

Utility-dependent improving directions at two steps. For the grid $(0, a, 1)$, maximizing the numerical utility is equivalent to maximizing $\ell_m(a) = \log(1 + a g_m(0)) + \log(1 + (1 - a)g_m(a))$, whose derivative is

$$\ell'_m(a) = \frac{g_m(0)}{1 + a g_m(0)} + \frac{(1 - a)g'_m(a) - g_m(a)}{1 + (1 - a)g_m(a)}. \tag{29}$$

Direct interval evaluation at the uniform node gives $\ell'_1(1/2) \in [-0.15929, -0.15927]$ and $\ell'_2(1/2) \in [0.01168, 0.01170]$. Differentiability therefore implies that a sufficiently small earlier movement improves u_1, whereas a later movement improves u_2. This proves utility-dependent improving directions at $K = 2$; it does not assert global finite-budget optimality of the analytical allocations.

Leading-order signed-error prediction. The right-endpoint adjoint in Eq. (13) gives

$$P_{m,K}(\rho) = \frac{x_{1,m}}{2}\sum_{k=0}^{K-1} e^{G_m(\tau_k^\rho)-G_m(\tau_{k+1}^\rho)} q_m(\tau_k^\rho) h_k^2,$$

$$u_m(\widehat{\boldsymbol{x}}_1^\rho) - u_m(\boldsymbol{x}_1) = -P_{m,K}(\rho) + o(K^{-1}).$$

Thus $P_{m,K}$ predicts $L_{m,K}$, whereas $-P_{m,K}$ predicts the signed utility change. Replacing the right-endpoint adjoint by its left-endpoint value removes the exponential factor and changes the sum by $O(K^{-2})$; we retain that factor in the numerical check. Report

$$\varepsilon_{m,K}(\rho) = 100\frac{|P_{m,K}(\rho) - L_{m,K}(\rho)|}{L_{m,K}(\rho)}, \qquad R_{m,K}(\rho) = 100\left(1 - \frac{L_{m,K}(\rho)}{L_{m,K}(1)}\right).$$

Positive $R_{m,K}$ means less terminal-utility error than uniform, not the same percentage increase in absolute utility. Both ratios cancel the source amplitude exactly. Evaluating at $\boldsymbol{x}_0 = (1, 1)$ also gives the expected absolute errors under the specified source distribution.

Numericals. Tab. 5 recalculates the exact-formula checks over $K \in \{2, 4, 8, 16, 32, 64, 128, 256\}$, using float64, adaptive one-dimensional quadrature (absolute and relative tolerances 10^{-12}), and linear inversion of the piecewise-linear CDF. Tightening both tolerances to 10^{-13} leaves every displayed value unchanged. The Euler product uses $\log(1 + z)$ and $\exp(z) - 1$ forms for stability. These calculations are independent of the separate trajectory/backward-adjoint integration check, which agreed with the exact solutions to approximately 1.23×10^{-10}.

The binned allocations reduce actual terminal-utility error for both utilities at every tested budget, including $K = 2, 4$, although the leading-order predictor is inaccurate there. By contrast, the full discrete-adjoint identity has zero analytical remainder; direct evaluation agrees with the endpoint loss to floating-point precision. At NFE 256, the maximum leading-order prediction discrepancy falls to 0.436%. Allocation benefit and asymptotic prediction accuracy are therefore separate checks.

Interval-enclosed low-budget improvements. To verify strict improvement without relying on quadrature error heuristics, we enclose each of the 64 bin integrals by sums of integrand ranges on a dyadic mesh of 65,536 cells over $[0, 1]$. Interval arithmetic at 35 decimal digits propagates these enclosures through the piecewise-linear CDF inverse and exact endpoint formulas. Outward-rounded percentage-reduction enclosures are

$$R_{1,2}(\rho_{1,B}) \in [14.736, 14.802], \quad R_{2,2}(\rho_{2,B}) \in [43.454, 43.492],$$
$$R_{1,4}(\rho_{1,B}) \in [19.779, 20.314], \quad R_{2,4}(\rho_{2,B}) \in [48.999, 49.207].$$

These strictly positive numerical enclosures concern the specified analytical allocations, not statistical uncertainty or globally optimal grids. Source-amplitude cancellation makes them valid for every

Table 5: Controlled Euler validation from the exact formulas. Leading-order prediction discrepancy is the maximum of $\varepsilon_{m,K}$ over uniform and the matching 64-bin adjoint allocation. Error reduction is $R_{m,K}(\rho_{m,B})$. All entries except NFE are percentages.

	Leading-order discrepancy ↓		Error reduction ↑	
NFE	u_1	u_2	u_1	u_2
2	162.984	92.429	14.77	43.47
4	32.638	23.505	20.05	49.10
8	7.263	1.767	25.16	50.31
16	1.116	2.204	26.07	49.84
32	0.171	2.121	26.20	49.58
64	0.237	1.418	26.18	49.33
128	0.165	0.813	26.13	49.17
256	0.095	0.436	26.10	49.07

draw in the stated source distribution. For continuous allocation, the asymptotic reduction follows from the surrogate:

$$\lim_{K\to\infty} R_{m,K}(\rho_m^\star) = 100\left[1 - \frac{\left(\int_0^1 \sqrt{q_m(t)}\,\mathrm{d}t\right)^2}{\int_0^1 q_m(t)\,\mathrm{d}t}\right],$$

namely 26.09% for u_1 and 49.01% for u_2. The finite-bin surrogate excess, the leading-order prediction discrepancy, and the exact finite-budget allocation benefit are distinct quantities. This example validates their relationship for smooth, sign-consistent Euler defects.

A.6 Prescribed Coordinates and Clock Non-Identifiability

Let $z \in C^1([0,1])$ be a strictly monotone coordinate on the finite inference interval, with $0 < c_z \le |z'(t)| \le C_z < \infty$. Uniform spacing in z induces the flow-time density

$$\rho_z(t) = \frac{|z'(t)|}{\int_0^1 |z'(a)|\,\mathrm{d}a}, \qquad F_{\rho_z}(t) = \frac{|z(t) - z(0)|}{|z(1) - z(0)|}. \tag{30}$$

Thus inverse-CDF placement under ρ_z is exactly uniform placement in the normalized z-coordinate.

Corollary 1 (Optimality condition for a prescribed coordinate)**.** *Under the assumptions of Theorem 2(ii), the coordinate-induced clock ρ_z is the unique minimizer of the allocation surrogate in Eq. (14) if and only if, for some constant $C_\xi > 0$,*

$$\kappa_\xi(t) = C_\xi |z'(t)|^{p_s+1} \qquad \textit{for almost every } t. \tag{31}$$

Proof. By Theorem 2, the unique unconstrained minimizer satisfies $\rho_\xi^\star(t) \propto \kappa_\xi(t)^{1/(p_s+1)}$. Comparing this expression with $\rho_z(t) \propto |z'(t)|$ gives Eq. (31), and the converse follows by substitution. □

For a diffusion probability-flow ODE, z may be a noise-scale or log-SNR coordinate inherited from the chosen forward corruption process. Corollary 1 characterizes when this reference clock minimizes the allocation surrogate: its rate must match the propagated task-cost profile. This condition does not establish true-utility optimality. For flow matching and rectified flow, path time is also a chosen parameterization rather than a coordinate intrinsic to the endpoint map: for any smooth increasing bijection $\vartheta : [0,1] \to [0,1]$,

$$\widetilde{\boldsymbol{v}}_\theta(\boldsymbol{x}, \lambda; \boldsymbol{c}) = \vartheta'(\lambda)\boldsymbol{v}_\theta(\boldsymbol{x}, \vartheta(\lambda); \boldsymbol{c})$$

induces the same exact endpoint transport, while uniform finite-step grids in t and λ generally produce different numerical endpoints. The endpoint map therefore does not determine a preferred inference coordinate, even when training specifies a path parameterization.

Scalar witness. Consider $\dot{x}_t = 2t$ with arbitrary x_0, so $x_1 = x_0 + 1$. For forward Euler and terminal utility $u(x) = x$, the effective profile is constant, $\kappa_\xi(t) \equiv 1$, and the surrogate optimum is the uniform clock. Now take the monotone coordinate $z(t) = \frac{3}{2}t - \frac{1}{2}t^2$, which induces $\rho_z(t) = \frac{3}{2} - t$. For $p_s = 1$,

$$\frac{\mathcal{A}_\xi(\rho_z)}{\mathcal{A}_\xi(1)} = \int_0^1 \frac{\mathrm{d}t}{\frac{3}{2} - t} = \log 3 > 1.$$

Thus even a smooth prescribed coordinate can be suboptimal for the surrogate. At $K = 2$, uniform flow-time Euler gives $\widehat{x}_1 = x_0 + \frac{1}{2}$, whereas uniform spacing in z uses the interior node $a = (3 - \sqrt{5})/2$ and gives $\widehat{x}_1 = x_0 + 2a(1 - a) \approx x_0 + 0.472$, although the exact endpoint is unchanged.

For any finite grid in this example, with $h_k = \tau_{k+1} - \tau_k$, $\widehat{x}_1 = x_0 + 1 - \sum_k h_k^2$. Since $\sum_k h_k = 1$, Cauchy–Schwarz gives $\sum_k h_k^2 \geq 1/K$, with equality only for uniform widths. Thus smoothness alone cannot guarantee that a nonuniform allocation improves finite-budget utility.

A.7 KL-REGULARIZED UTILITY TILT

The whole-clock decision in Sec. 3 is a one-step RL problem: ξ is the observed setting, ρ the action, and $S_\xi(\rho)$ its expected terminal utility. Solver steps execute the chosen action rather than supply further policy decisions. Fitting this utility from fixed completed-solve evidence therefore requires neither a transition model nor bootstrapped value targets. The variational problem below optimizes a law over clock actions; Appx. A.10 relates it to the deterministic allocation policy.

Theorem 3 (Gibbs variational characterization)**.** *Fix ξ, let $q_0(\cdot \mid \xi)$ be a probability law on $\mathcal{R}_\mathrm{f}$, assume S_ξ is measurable, and let $0 < Z_\beta(\xi) < \infty$. If $\pi^\star_\beta$ has finite expected utility and finite KL—as holds for bounded scores and finite candidate laws with finite scores—then Eq. (8) uniquely maximizes Eq. (7) over $\pi \ll q_0$ for which both terms are finite.*

Proof. Suppressing ξ,

$$\frac{\mathrm{d}\pi^\star_\beta}{\mathrm{d}q_0}(\rho) = \frac{e^{\beta S_\xi(\rho)}}{Z_\beta(\xi)}.$$

Hence, for any admissible $\pi \ll q_0$,

$$D_{\mathrm{KL}}(\pi \| \pi^\star_\beta) = D_{\mathrm{KL}}(\pi \| q_0) - \beta \mathbb{E}_\pi[S_\xi] + \log Z_\beta(\xi).$$

Rearranging gives

$$\mathbb{E}_\pi[S_\xi] - \frac{1}{\beta} D_{\mathrm{KL}}(\pi \| q_0) = \frac{1}{\beta} \log Z_\beta(\xi) - \frac{1}{\beta} D_{\mathrm{KL}}(\pi \| \pi^\star_\beta),$$

whose unique maximizer is $\pi = \pi^\star_\beta$ almost everywhere. □

If $\widetilde{S}_\xi(\rho) = c_1 S_\xi(\rho) + c_0$ with $c_1 > 0$, then the tilt built from $(\widetilde{S}_\xi, \beta)$ equals that built from $(S_\xi, c_1\beta)$: the offset cancels and the scale changes the effective temperature. A nonlinear monotone transformation need not admit this reduction and can therefore change the tilted law.

A.8 ALLOCATION-SURROGATE SPECIALIZATION

Suppose $S_\xi(\rho) = S_{\xi,0} - \mathcal{A}_\xi(\rho)$, and define the support-accessible optimum

$$\mathcal{A}^\star_{\xi,q_0} = \operatorname*{ess\,inf}_{\rho \sim q_0(\cdot|\xi)} \mathcal{A}_\xi(\rho) < \infty.$$

For $\eta > 0$, let

$$\mathcal{N}_{\xi,\eta} = \{\rho : \mathcal{A}_\xi(\rho) \leq \mathcal{A}^\star_{\xi,q_0} + \eta\}.$$

Then $q_0(\mathcal{N}_{\xi,\eta} \mid \xi) > 0$ by the definition of essential infimum.

Corollary 2 (Support-limited concentration)**.** *For $0 < \eta < \varepsilon$,*

$$\pi^\star_\beta\big(\mathcal{A}_\xi(\rho) > \mathcal{A}^\star_{\xi,q_0} + \varepsilon \mid \xi\big) \leq \frac{e^{-\beta(\varepsilon-\eta)}}{q_0(\mathcal{N}_{\xi,\eta} \mid \xi)}. \tag{32}$$

Proof. The clock-independent baseline cancels, so the tilted density is proportional to $e^{-\beta \mathcal{A}_\xi(\rho)}$. On the event in Eq. (32), the unnormalized mass is at most $e^{-\beta(\mathcal{A}^\star_{\xi,q_0}+\varepsilon)}$. The normalizer is at least its contribution on $\mathcal{N}_{\xi,\eta}$, namely $q_0(\mathcal{N}_{\xi,\eta} \mid \xi)e^{-\beta(\mathcal{A}^\star_{\xi,q_0}+\eta)}$. Taking the ratio proves the bound. □

For fixed ξ, increasing β concentrates only toward support-accessible near-minimizers, not optima outside that support. If the solver, profile, and base law are fixed across K, the exponent depends on βK^{-p_s}; keeping this product constant preserves the tilt. This does not prescribe a temperature rule for GICO's measured utilities.

A.9 Sequential Doob h-transform

We apply the classical Doob h-transform (Doob, 1957; Chetrite & Touchette, 2015; Léonard, 2014) to clock-construction paths, not feedback decisions along the ODE solve. Fix ξ. Let an initial construction state z_0 and Markov transitions $Q_{\xi,k}(\mathrm{d}\chi_k \mid \mathsf{z}_k)$, with $\mathsf{z}_{k+1} = \mathsf{T}_k(\mathsf{z}_k, \chi_k)$, induce a path law Q_ξ. The terminal state determines $\rho(\mathsf{z}_L)$, and its pushforward under Q_ξ is $q_0(\cdot \mid \xi)$. Set

$$\Gamma_\xi(\mathsf{z}_L) = e^{\beta S_\xi(\rho(\mathsf{z}_L))}, \qquad h_{\xi,k}(\mathsf{z}_k) = \mathbb{E}_{Q_\xi}[\Gamma_\xi(\mathsf{z}_L) \mid \mathsf{z}_k], \tag{33}$$

and assume $0 < h_{\xi,k}(\mathsf{z}_k) < \infty$ on the relevant states.

Theorem 4 (Doob realization of the terminal tilt)**.** *The kernels*

$$Q^h_{\xi,k}(\mathrm{d}\chi_k \mid \mathsf{z}_k) = Q_{\xi,k}(\mathrm{d}\chi_k \mid \mathsf{z}_k)\frac{h_{\xi,k+1}(\mathsf{T}_k(\mathsf{z}_k,\chi_k))}{h_{\xi,k}(\mathsf{z}_k)} \tag{34}$$

are normalized Markov transitions. Their path law Q^h_ξ satisfies

$$\frac{\mathrm{d}Q^h_\xi}{\mathrm{d}Q_\xi} = \frac{\Gamma_\xi(\mathsf{z}_L)}{h_{\xi,0}(\mathsf{z}_0)}, \qquad h_{\xi,0}(\mathsf{z}_0) = Z_\beta(\xi),$$

and its terminal clock marginal is exactly $\pi^\star_\beta(\cdot \mid \xi)$.

Proof. The tower property gives

$$\int Q_{\xi,k}(\mathrm{d}\chi_k \mid \mathsf{z}_k)h_{\xi,k+1}(\mathsf{T}_k(\mathsf{z}_k,\chi_k)) = h_{\xi,k}(\mathsf{z}_k),$$

so each kernel is normalized. Along a completed path, the likelihood ratios telescope:

$$\prod_{k=0}^{L-1}\frac{h_{\xi,k+1}(\mathsf{z}_{k+1})}{h_{\xi,k}(\mathsf{z}_k)} = \frac{h_{\xi,L}(\mathsf{z}_L)}{h_{\xi,0}(\mathsf{z}_0)} = \frac{\Gamma_\xi(\mathsf{z}_L)}{h_{\xi,0}(\mathsf{z}_0)}.$$

Since z_0 is fixed and the clock pushforward of Q_ξ is $q_0(\cdot \mid \xi)$,

$$h_{\xi,0}(\mathsf{z}_0) = \mathbb{E}_{Q_\xi}[e^{\beta S_\xi(\rho(\mathsf{z}_L))}] = Z_\beta(\xi).$$

For every measurable clock set A,

$$Q^h_\xi(\rho(\mathsf{z}_L) \in A) = \frac{\int_A e^{\beta S_\xi(\rho)} q_0(\mathrm{d}\rho \mid \xi)}{Z_\beta(\xi)} = \pi^\star_\beta(A \mid \xi).$$

□

The potential uses expected utility $S_\xi(\rho)$. If a completed evaluation Y remains random conditional on ρ, tilting an augmented path by $e^{\beta Y}$ instead induces

$$\pi(\mathrm{d}\rho) \propto \mathbb{E}[e^{\beta Y} \mid \rho, \xi]q_0(\mathrm{d}\rho \mid \xi),$$

which generally differs from $e^{\beta \mathbb{E}[Y \mid \rho,\xi]}q_0(\mathrm{d}\rho \mid \xi)$. The two coincide when Y is deterministic conditional on the completed clock and setting, but not in general.

A.10 FINITE-SUPPORT TILT AND CENTROID-ANCHORED REFINEMENT

For the indexed candidate family $\mathcal{C}_\xi = (\rho_i)_{i=1}^{n_\xi}$, define

$$q_{0,n_\xi}(\mathrm{d}\rho \mid \xi) = \sum_{i=1}^{n_\xi} w_i^0(\xi)\delta_{\rho_i}(\mathrm{d}\rho), \qquad w_i^0(\xi) > 0, \quad \sum_i w_i^0(\xi) = 1.$$

This is a configured design law, not a behavior policy estimated from collection frequencies. Experiments use the fixed reference pool in Appx B.2; positive base weights are specified before utility evaluation and never updated from locked-test outcomes. Let $\Delta_{n_\xi} = \{\boldsymbol{w} \in \mathbb{R}^{n_\xi} : w_i \geq 0,\ \sum_i w_i = 1\}$. Treating the measured scores as fixed, specializing Theorem 3 to this support gives

$$w_i^\star(\xi) = \frac{w_i^0(\xi)e^{\beta\widehat{S}_i(\xi)}}{\sum_{j=1}^{n_\xi} w_j^0(\xi)e^{\beta\widehat{S}_j(\xi)}}, \tag{35}$$

which uniquely solves

$$\max_{\boldsymbol{w}\in\Delta_{n_\xi}} \left\{ \sum_i w_i \widehat{S}_i(\xi) - \frac{1}{\beta}\sum_i w_i \log\frac{w_i}{w_i^0(\xi)} \right\}. \tag{36}$$

The surrogate-defined weights in Eq. (11) replace $\widehat{S}_i(\xi)$ by $\bar{r}_\omega(\rho_i, \xi)$ and β by $1/T_{\mathrm{tilt}}$. If $\bar{r}_\omega(\rho_i, \xi) = c_1\widehat{S}_i(\xi) + c_0$ exactly with $c_1 > 0$, the effective inverse temperature is c_1/T_{tilt}, while c_0 cancels. If the maximum surrogate score is unique, $T_{\mathrm{tilt}} \downarrow 0$ concentrates on that candidate; under ties, the base-weight proportions among maximizers are retained.

Lemma 4 (Forward-KL centroid)**.** *Let $\boldsymbol{p}_i \in \Delta_B$, $w_i \geq 0$, $\sum_i w_i = 1$, and $\bar{\boldsymbol{p}} = \sum_i w_i\boldsymbol{p}_i$. Then, for every $\boldsymbol{p} \in \Delta_B$, with the standard extended-value convention for KL,*

$$\sum_i w_i D_{\mathrm{KL}}(\boldsymbol{p}_i\|\boldsymbol{p}) = \sum_i w_i D_{\mathrm{KL}}(\boldsymbol{p}_i\|\bar{\boldsymbol{p}}) + D_{\mathrm{KL}}(\bar{\boldsymbol{p}}\|\boldsymbol{p}). \tag{37}$$

Consequently, $\bar{\boldsymbol{p}}$ is the unique minimizer of the weighted forward-KL objective.

Proof. Subtracting the first term on the right from the left gives

$$\sum_{i,j} w_i p_{ij}\log\frac{\bar{p}_j}{p_j} = \sum_j \bar{p}_j \log\frac{\bar{p}_j}{p_j} = D_{\mathrm{KL}}(\bar{\boldsymbol{p}}\|\boldsymbol{p}).$$

Nonnegativity of KL gives the result. □

Define $p_j(\rho) = \int_{j/B}^{(j+1)/B} \rho(t)\,\mathrm{d}t$, $j = 0, \ldots, B-1$. Since the terminal clock marginal of Q_ξ^h is $\pi_\beta^\star(\cdot \mid \xi)$, Lemma 4 gives the population projection

$$\boldsymbol{p}_\beta^{\mathrm{cent}}(\xi) = \operatorname*{arg\,min}_{\boldsymbol{p}\in\Delta_B} \mathbb{E}_{\rho\sim\pi_\beta^\star(\cdot|\xi)} D_{\mathrm{KL}}(\boldsymbol{p}(\rho)\|\boldsymbol{p}) = \mathbb{E}_{Q_\xi^h}[\boldsymbol{p}(\rho(\mathsf{z}_L))]. \tag{38}$$

Eq. (11) replaces the population law by the finite surrogate-induced tilt $\widehat{\pi}_\omega(\cdot \mid \xi)$. Its centroid is one allocation action, not a distribution over clock actions. The weighted forward-KL fit is therefore an allocation projection, not whole-clock likelihood fitting. It supplies the anchor, not the optimizer of the score-augmented objective or a reconstruction of the Doob kernels. For the fixed-bin reconstruction, $\rho_{\bar{\boldsymbol{p}}} = \sum_i w_i\rho_{\boldsymbol{p}_i}$, and pointwise convexity of $u \mapsto u^{-p_s}$ gives

$$\mathcal{A}_\xi(\rho_{\bar{\boldsymbol{p}}}) \leq \sum_i w_i \mathcal{A}_\xi(\rho_{\boldsymbol{p}_i}). \tag{39}$$

This Jensen bound compares the anchor's surrogate cost with the weighted mean reference cost. It neither proves improvement over the best candidate nor bounds the true terminal utility. The bound holds for the centroid of guarded references, but need not hold after policy refinement.

KL-anchored refinement. Fix a cell with positive anchor $\bar{\boldsymbol{p}} = \bar{\boldsymbol{p}}_\xi$, set $\eta = \eta_{\mathrm{aux}} \geq 0$, and write

$$J_\eta(\boldsymbol{p}) = D_{\mathrm{KL}}(\bar{\boldsymbol{p}}\|\boldsymbol{p}) - \eta C_\omega(\boldsymbol{p}, \xi).$$

This objective optimizes a clock action against a fitted utility model, penalizing displacement in allocation space. It is not the clock-law objective in Eq. (7). Any feasible $\boldsymbol{p}_\eta$ satisfying $J_\eta(\boldsymbol{p}_\eta) \le J_\eta(\bar{\boldsymbol{p}})$ obeys

$$D_{\mathrm{KL}}(\bar{\boldsymbol{p}} \| \boldsymbol{p}_\eta) \le \eta\{C_\omega(\boldsymbol{p}_\eta, \xi) - C_\omega(\bar{\boldsymbol{p}}, \xi)\} \le 10\eta. \tag{40}$$

The first inequality rearranges the objective comparison; the second uses the clip range. For $\eta > 0$, such a refinement cannot lower the clipped surrogate utility. At $\eta = 0$, the unique optimum is the anchor. Positive η allows departures, including densities outside the candidates' convex hull, but penalizes them according to forward KL.

The local direction is explicit. If C_ω is twice continuously differentiable near $\bar{\boldsymbol{p}}$, the interior local minimizer continuous from $\boldsymbol{p}_0 = \bar{\boldsymbol{p}}$ satisfies, as $\eta \downarrow 0$,

$$\boldsymbol{p}_\eta = \bar{\boldsymbol{p}} + \eta\big[\mathrm{diag}(\bar{\boldsymbol{p}}) - \bar{\boldsymbol{p}}\bar{\boldsymbol{p}}^\top\big]\boldsymbol{g} + O(\eta^2), \qquad \boldsymbol{g} = \nabla_{\boldsymbol{p}} C_\omega(\bar{\boldsymbol{p}}, \xi). \tag{41}$$

Indeed, stationarity under $\sum_j p_j = 1$ gives $-\bar{p}_j/p_j - \eta\,\partial_j C_\omega + \lambda_{\mathrm{KKT}} = 0$. Expanding about $(\boldsymbol{p}, \lambda_{\mathrm{KKT}}) = (\bar{\boldsymbol{p}}, 1)$ yields $p_j - \bar{p}_j = \eta\bar{p}_j(g_j - \bar{\boldsymbol{p}}^\top\boldsymbol{g}) + O(\eta^2)$. This is the natural-gradient direction for the categorical Fisher metric $g_{\bar{\boldsymbol{p}}}(\boldsymbol{d}, \boldsymbol{d}) = \sum_j d_j^2/\bar{p}_j$ on the tangent space $\sum_j d_j = 0$. It increases predicted utility to first order by $\eta\{\sum_j \bar{p}_j g_j^2 - (\sum_j \bar{p}_j g_j)^2\}$, a nonnegative weighted variance. It describes the per-cell objective's local solution, not the trajectory of AdamW or an exact projection of the full clock law.

Conditional true-utility improvement at a fixed budget. The objective comparison can be transferred to true utility when calibration error is controlled at the two compared allocations. Let $S(\boldsymbol{p}) = S_\xi(\rho_{\boldsymbol{p}})$ in the fixed utility units fitted by $\bar{r}_\omega$, and define

$$\widetilde{S}_\omega(\boldsymbol{p}) = \mu_\xi + \sigma_\xi C_\omega(\boldsymbol{p}, \xi), \qquad \sigma_\xi > 0.$$

The fixed μ_ξ, σ_ξ are those in Appx. B.4. Thus $\widetilde{S}_\omega$ is the clipped refinement score mapped back to these utility units, not the unclipped surrogate prediction.

Proposition 3 (Calibrated finite-budget refinement)**.** *Fix $\eta > 0$ and a feasible $\boldsymbol{p}$ satisfying $J_\eta(\boldsymbol{p}) \le J_\eta(\bar{\boldsymbol{p}}) + \delta_{\mathrm{opt}}$, where $\delta_{\mathrm{opt}} \ge 0$. If $|\widetilde{S}_\omega(\boldsymbol{p}) - S(\boldsymbol{p})| \le \epsilon_{\boldsymbol{p}}$ and $|\widetilde{S}_\omega(\bar{\boldsymbol{p}}) - S(\bar{\boldsymbol{p}})| \le \epsilon_{\bar{\boldsymbol{p}}}$, then*

$$S(\boldsymbol{p}) - S(\bar{\boldsymbol{p}}) \ge \frac{\sigma_\xi}{\eta}\left[D_{\mathrm{KL}}(\bar{\boldsymbol{p}} \| \boldsymbol{p}) - \delta_{\mathrm{opt}}\right] - \epsilon_{\boldsymbol{p}} - \epsilon_{\bar{\boldsymbol{p}}}. \tag{42}$$

A strictly positive lower bound certifies improvement over the anchor, without an NFE asymptotic or a differentiable terminal scorer.

Proof. The objective comparison gives $\eta\{C_\omega(\boldsymbol{p}, \xi) - C_\omega(\bar{\boldsymbol{p}}, \xi)\} \ge D_{\mathrm{KL}}(\bar{\boldsymbol{p}} \| \boldsymbol{p}) - \delta_{\mathrm{opt}}$. Multiplication by σ_ξ/η and subtraction of the two calibration error bounds prove the claim. □

The errors include clipping bias and must hold at the anchor and refined action, not just at the reference clocks. This comparison is with the anchor, not necessarily with uniform. A shared fitted policy need not satisfy the per-setting objective comparison, and allocation KL neither enforces logged-action support nor establishes these accuracy bounds. Clipping, delayed refinement, and selection by held-out anchor KL and frozen-surrogate utility limit the optimization incentive (Appx. C.2); they do not supply the calibration certificate required by Proposition 3.

A.11 Stochastic clock policies and causal distillation

Unlike the deterministic allocation policy, this extension fits a law over whole-clock constructions. Its autoregressive coordinates are sampled before integration, not from intermediate solver states. The population likelihood anchor follows Theorem 4. Let $\boldsymbol{\chi} = (\chi_0, \ldots, \chi_{L-1})$ denote a complete construction path and factor

$$P_\psi(\mathrm{d}\boldsymbol{\chi} \mid \xi) = \prod_{k=0}^{L-1} P_{\psi,k}(\mathrm{d}\chi_k \mid \mathsf{z}_k, \xi). \tag{43}$$

Assume these kernels admit conditional densities $p_{\psi,k}$ with respect to the same reference measures as the base kernels. From Theorem 4,

$$W_\xi(\boldsymbol{\chi}) := \frac{e^{\beta S_\xi(\rho(\mathsf{z}_L))}}{Z_\beta(\xi)} = \frac{\mathrm{d}Q_\xi^h}{\mathrm{d}Q_\xi}(\boldsymbol{\chi}).$$

The likelihood-only population objective

$$\mathcal{L}_{\mathrm{causal}}(\psi) = -\mathbb{E}_{\xi}\mathbb{E}_{\boldsymbol{\chi}\sim Q_{\xi}}\left[W_{\xi}(\boldsymbol{\chi})\sum_{k=0}^{L-1}\log p_{\psi,k}(\chi_k \mid \mathsf{z}_k, \xi)\right] \tag{44}$$

differs from $\mathbb{E}_{\xi} D_{\mathrm{KL}}(Q^h_{\xi} \| P_{\psi}(\cdot \mid \xi))$ only by a ψ-independent term whenever the quantities are finite. Hence an expressive optimum recovers the Doob-transformed conditionals on Q^h_{ξ}-supported states without explicitly regressing the $h_{\xi,k}$-functions.

For the empirical implementation, encode each guarded reference density as a standardized log-ratio vector $\boldsymbol{z}_i \in \mathbb{R}^L$, with $L = B - 1 = 63$, using Appx B.5. With fixed surrogate weights from Eq. (11), define the smoothed anchor distribution

$$\widehat{Q}_{\xi,\epsilon} = \sum_{i=1}^{n_{\xi}} \bar{w}_i(\xi)\mathcal{N}(\boldsymbol{z}_i, \epsilon^2 \boldsymbol{I}_L), \qquad \epsilon = 0.1.$$

Each continuous coordinate $\chi_k = z_k$ is conditioned on the prefix $\boldsymbol{\chi}_{<k}$; $\boldsymbol{p}(\boldsymbol{\chi})$ denotes the decoded, guarded masses. With Gaussian conditional densities relative to Lebesgue measure, the implemented objective is

$$\begin{aligned} \widehat{\mathcal{L}}_{\mathrm{causal}}(\psi) &= -\mathbb{E}_{\xi}\mathbb{E}_{\boldsymbol{\chi}\sim\widehat{Q}_{\xi,\epsilon}}\sum_{k=0}^{L-1}\log p_{\psi,k}(\chi_k \mid \boldsymbol{\chi}_{<k}, \xi), \\ \mathcal{L}_{\mathrm{sto}}(\psi) &= L^{-1}\widehat{\mathcal{L}}_{\mathrm{causal}}(\psi) - \eta_{\mathrm{aux}}\mathbb{E}_{\xi}\mathbb{E}_{\boldsymbol{\chi}\sim P_{\psi}(\cdot|\xi)} C_{\omega}(\boldsymbol{p}(\boldsymbol{\chi}), \xi). \end{aligned} \tag{45}$$

The first term equals $L^{-1}\mathbb{E}_{\xi} D_{\mathrm{KL}}(\widehat{Q}_{\xi,\epsilon} \| P_{\psi})$ plus a parameter-independent term when finite. Smoothing supplies a continuous empirical target: an atomic reference law would not have finite forward KL to this continuous policy. This target is not the population Doob law Q^h_{ξ}; finite evidence, surrogate approximation, smoothing, and the Gaussian conditional family do not guarantee recovery of that law. Auxiliary refinement uses fresh reparameterized clock samples scored only by the frozen surrogate (Appx B.4). These are model-based proposals, not newly scored generator rollouts; fitting remains offline and does not guarantee true-utility gains.

A clock law and its mean density need not yield the same numerical outputs: both inverse-CDF placement and finite-step integration are nonlinear. Expected per-clock utility is linear in the clock law and bounded by its support supremum. A nonlinear metric of the mixed endpoint distribution need not obey this ordering. Therefore, expected per-clock utility and quality measured after pooling outputs from different clocks are distinct.

B Implementation

This appendix specifies clock construction and the implemented objectives. Architecture and fitting are in Appx C; data and evaluation are in Appx D.

B.1 Finite density representation and decoding

Use $B = 64$ uniform bins with edges $e_j = j/B$. Raw masses $\boldsymbol{a} \in \Delta_B$ are mapped to positive masses by

$$G_{\delta}(\boldsymbol{a}) = (1-\delta)\boldsymbol{a} + \delta \mathbf{1}/B, \qquad \delta = 10^{-8}, \qquad \boldsymbol{p} = G_{\delta}(\boldsymbol{a}).$$

Surrogate log-mass features, density KL, and deployed CDFs use $\boldsymbol{p}$, whose density is

$$\rho_{\boldsymbol{p}}(t) = \frac{p_j}{e_{j+1} - e_j}, \qquad t \in [e_j, e_{j+1}). \tag{46}$$

The affine map G_{δ} commutes with weighted averaging, so the reference centroid and deployed densities use the same representation.

For a reference grid $0 = \tau_0 < \cdots < \tau_K = 1$, assigning equal clock mass to each interval gives

$$a_j = \sum_{k=0}^{K-1} \frac{|[e_j, e_{j+1}) \cap [\tau_k, \tau_{k+1})|}{K(\tau_{k+1} - \tau_k)}. \tag{47}$$

The float64 CDF is inverted by linear interpolation at k/K, with endpoints pinned to zero and one. The generic decoder repairs a non-increasing node by a 10^{-8} increment, capped at one, then validates the grid; deployment also requires strict ordering in float32. Masses must be finite, nonnegative, and normalized.

Reference utilities are measured on the encoded-and-decoded grids, which can differ from the source grids. Each saved observation records the raw masses and exact executed nodes, together with its context, solver, realized NFE, and frozen backbone.

B.2 Reference-clock mapping and solver/NFE semantics

The pool contains 25 named clocks: uniform; native-coordinate and log-σ AYS, GITS, and OTS; Late-p for $p \in \{1.5, 2, 3, 4, 8\}$; FlowTS with exponent 0.03; and the reversals of the 12 nonuniform clocks. Identical realized densities are merged before weighting, with a uniform prior over distinct densities.

For a descending native coordinate $z_0 > \cdots > z_J$, transfer to normalized flow time uses

$$t_j = \frac{z_0 - z_j}{z_0 - z_J}.$$

AYS uses the listed SD1.5 timesteps or log-σ nodes; its terminal positive σ is reconstructed from the pinned SD1.5 scaled-linear beta schedule. GITS uses the published CIFAR-10 schedule exposed by the official implementation (Chen et al., 2024),

$$\boldsymbol{\sigma} = (80,\ 10.9836,\ 3.8811,\ 1.5840,\ 0.5666,\ 0.1698,\ 0.0020).$$

AYS and GITS are linearly interpolated over normalized source-node index, not directly over σ or time. OTS grids are computed once with the official step-optimization routine (Xue et al., 2024), using the continuous linear-VP schedule and uniform-time initialization at a fixed source revision. They are stored for macro-step counts $K \in \{2, 3, 4, 5, 6, 7, 8, 10, 12, 14, 16, 20\}$. OTS is evaluated only at these stored counts.

For $\lambda_j = j/K$, the analytic reference grids and the reversal of a base grid $\{\tau_j\}_{j=0}^{K}$ are

$$\tau_j^{\text{late-}p} = 1 - (1 - \lambda_j)^p, \qquad \tau_j^{\text{FlowTS}} = \lambda_j^{q_{\text{F}}}, \quad q_{\text{F}} = 0.03, \qquad \tau_j^{\text{rev}} = 1 - \tau_{K-j}.$$

Late-p gives polynomial late allocation. The FlowTS-inspired power clock follows t-power sampling (Hu et al., 2024), with $q_{\text{F}} = 0.03$ chosen for this benchmark. Before the uniform mixture, their continuum densities on $(0, 1)$ are

$$\rho_{\text{late-}p}(t) = \frac{1}{p}(1 - t)^{1/p - 1}, \qquad \rho_{\text{FlowTS}}(t) = \frac{1}{q_{\text{F}}} t^{1/q_{\text{F}} - 1}.$$

Reference grids use Eq. (47), subject to the representation check in Appx B.1.

Table 6: Solver and NFE semantics. In the paper, b is the realized number of frozen-field evaluations and $K_s(b)$ is the macro-step count.

Solver	$K_s(b)$	Evaluations per macro-step	Startup / fixed rule
Euler	b	1	Forward Euler.
Heun	$b/2$, even b	2	Predictor–corrector.
Midpoint RK2	$b/2$, even b	2	Explicit midpoint.
Variable-step AB2	b	1	First step Euler; subsequent two-step velocity update below.

Reported NFE is the field-evaluation count b; $K_s(b)$ is recorded separately. For variable-step Adams–Bashforth two-step (AB2), $k \geq 1$,

$$\widehat{\boldsymbol{x}}^{\rho}_{\tau^{\rho}_{k+1}} = \widehat{\boldsymbol{x}}^{\rho}_{\tau^{\rho}_{k}} + \Delta t^{\rho}_{k} \left[\left(1 + \frac{r^{\rho}_{k}}{2} \right) \boldsymbol{v}^{\rho}_{k} - \frac{r^{\rho}_{k}}{2} \boldsymbol{v}^{\rho}_{k-1} \right], \qquad r^{\rho}_{k} = \frac{\Delta t^{\rho}_{k}}{\Delta t^{\rho}_{k-1}}, \qquad \boldsymbol{v}^{\rho}_{k} := \boldsymbol{v}_{\theta}(\widehat{\boldsymbol{x}}^{\rho}_{\tau^{\rho}_{k}}, \tau^{\rho}_{k}; \boldsymbol{c}).$$

Euler startup and one reused velocity per subsequent step give $b = K$. COCO solvers and the frozen two-grid BézierFlow composition are specified in Appx G.1 and E.6.

B.3 MATCHED CELLS AND SURROGATE OBJECTIVE

A comparison group fixes context, backbone, solver, NFE, and measurement protocol. Collected candidates are paired with uniform using the same generation seeds, ensemble size, and reference data. The saved measurements remain fixed during learning. Repeats are averaged before nonlinear reward construction; batch metrics such as KID use complete feature blocks. Appx D.2 defines the calibrated utility components.

The surrogate predicts this utility vector. Within group ξ, let $\mathcal{P}_\xi$ contain measured reference pairs with distinct scalar utilities and set $\varsigma_{ij} = \operatorname{sign}(\widehat{S}_i - \widehat{S}_j)$. Ranking uses

$$\mathcal{L}_{\mathrm{rank}} = \mathbb{E}_\xi \left[\frac{1}{|\mathcal{P}_\xi|} \sum_{(i,j)\in\mathcal{P}_\xi} \operatorname{softplus}\left(-\frac{\varsigma_{ij}\{\bar{r}_\omega(\rho_i,\xi) - \bar{r}_\omega(\rho_j,\xi)\}}{T_{\mathrm{rank}}} \right) \right].$$

An empty pair set contributes zero. Component-weighted Huber regression (Huber, 1964) is averaged over measured references within each group; groups receive equal weight in both losses. Loss constants are in Tab. 10.

B.4 POLICY ANCHORS AND REFINEMENT

The frozen surrogate scores the complete distinct reference pool at each context/setting, including unmeasured combinations. These predictions add no observed utilities. Eq. (11) gives the weights and density centroid. For this fixed allocation target, cross entropy and forward KL have identical gradients:

$$\mathcal{L}_{\mathrm{CE}} = -\mathbb{E}_\xi \sum_j \bar{p}_{\xi,j} \log p_{\psi,j}(\xi) = \mathbb{E}_\xi D_{\mathrm{KL}}(\bar{\boldsymbol{p}}_\xi \| \boldsymbol{p}_\psi(\xi)) + \mathbb{E}_\xi H(\bar{\boldsymbol{p}}_\xi). \tag{48}$$

The stochastic anchor is the Gaussian-smoothed reference mixture in Appx A.11; it is fitted by continuous conditional likelihood. Both policies refine against the same frozen surrogate score:

$$C_\omega(\boldsymbol{p},\xi) = \operatorname{clip}\left(\frac{\bar{r}_\omega(\rho_{\boldsymbol{p}},\xi) - \mu_\xi}{\sigma_\xi}, -5, 5 \right). \tag{49}$$

Here μ_ξ and σ_ξ are the mean and population SD of surrogate predictions over the distinct references in that cell, frozen when targets are built; an SD below 10^{-6} is replaced by one. This clipping and standardization apply only to auxiliary refinement, not to the reward labels or mixture weights. Deterministic refinement optimizes the predicted allocation; stochastic refinement averages predicted utility over reparameterized clock draws. In both cases gradients pass through density inputs to the fixed surrogate, not through completed solves. Appx C.2 specifies the coefficient schedule and sampling counts.

B.5 CAUSAL DECODING AND DEPLOYMENT

For guarded masses $\boldsymbol{p}$, encode the first $L = B - 1 = 63$ log ratios relative to the last bin as

$$z_j = \frac{\log(p_j/p_{B-1}) - \mu_j^{\mathrm{z}}}{\sigma_j^{\mathrm{z}}}, \qquad j = 0, \ldots, B-2.$$

The coordinate means μ_j^{z} and population SDs σ_j^{z} are fitted on distinct surrogate-training reference densities and then frozen; SDs below 10^{-6} are replaced by one. The causal Transformer factorizes a continuous law into Gaussian conditionals

$$p_\psi(\boldsymbol{z} \mid \xi) = \prod_{j=0}^{L-1} \mathcal{N}\big(z_j; \mu_{\psi,j}(\boldsymbol{z}_{<j},\xi), \sigma^2_{\psi,j}(\boldsymbol{z}_{<j},\xi)\big), \quad \sigma_{\psi,j} = 0.05 + 1.95\,\mathrm{sigmoid}(h_{\psi,j}).$$

Decode by softmax of $(\mu_j^{\mathrm{z}} + \sigma_j^{\mathrm{z}} z_j)_{j<B-1}$ concatenated with a zero, followed by G_δ. The uniform mixture is part of the deployed map, so encoding and decoding need not be exact inverses. Appx A.11 defines the smoothed likelihood target.

A separate, replayable clock RNG supplies the 63 Gaussian innovations. One clock is sampled per generated image or trajectory before integration, independently of generation noise. Molecular policies condition on the initial observed history and reuse each member's clock throughout its rollout. Deployment performs no surrogate scoring, rejection, or reward-based clock selection.

B.6 METHOD SCOPE, SUPERVISION, AND DEPLOYMENT

Tab. 7 summarizes published formulations and the clock-PG and GICO-TF variants evaluated here. Task-specific baseline adaptations are in Appx G.3 and G.8.

Table 7: Sampling methods: optimized objects, fitting criteria, supervision, and deployment.

Method	Optimized object	Fitting / allocation criterion	Supervision	Deployment
BézierFlow	Continuous Bézier scheduler and evaluation times	Teacher-output matching; LPIPS for images	Accurate solves of the same frozen field; LPIPS features	Learned trajectory transform
SharpEuler	Euler density and grid	Acceleration-based power-law density; square-root special case	Field velocity differences along calibration trajectories	Fixed grid after offline calibration
LD3	Step and decoupled evaluation times	Soft teacher forcing; bounded training-noise adjustment; LPIPS for images	High-accuracy generated targets; perceptual features	Learned grid with ordinary generation noise
JS-REINFORCE	Prompt/noise-conditioned joint schedule	Terminal reward via REINFORCE with a James–Stein baseline	Scored generator rollouts	Whole schedule constructed before integration
Clock PG (this study)	Conditional whole-clock policy	PPO-style terminal-reward objective with KL and entropy terms	Scored generator rollouts	Whole clock selected before integration
GICO-det. / GICO-stoch.	Conditional density or whole-clock law	Centroid KL or smoothed likelihood, plus predicted-utility refinement	Saved utilities (offline); feature/scorer models as required by the utility	Predict or causally sample a clock
GICO-TF	GICO clock allocation	GICO objective with LPIPS target-fidelity utility	Generated targets and LPIPS features	Deploy the fitted clock

Teacher-output matching supplies generated targets; GICO's utility surrogate predicts measured terminal improvements. GICO-TF introduces LPIPS target fidelity (Appx E.5). Related methods retain distinct freedoms: S4S also learns solver coefficients, while RaSS and TPDM adapt to intermediate states rather than constructing an entire clock before integration (Frankel et al., 2025; Ding et al., 2025; Ye et al., 2025).

C ARCHITECTURES AND TRAINING

Each task/backbone has separately fitted weights using the shared architecture below. The offline protocol fixes collected evidence before fitting: surrogate selection uses saved utilities, and policy selection uses frozen-surrogate predictions. Neither fitting nor checkpoint selection calls a generator or terminal scorer.

C.1 MODEL ARCHITECTURES

Table 8: Shared GICO architecture.

Component	Specification
Shared blocks	Width 128; two pre-normalized Transformer layers; four attention heads; feed-forward width 256 with SiLU. RoPE at fixed bin centers and additive MLP conditioning (Vaswani et al., 2017; Su et al., 2024).
Utility surrogate	Tokens contain guarded log masses, bin centers, and bin widths. Unmasked bin attention and mean pooling predict the utility-component vector.
Deterministic policy	Bin-center/width query tokens and unmasked attention predict 64 logits; softmax followed by the common guard gives the deployed masses.
Causal policy	Shifted continuous log-ratio inputs and causal attention with RoPE predict 63 Gaussian conditionals (Appx B.5).

Table 9: Native conditioning and task-specific fitting settings.

Domain	Native context	Dropout	Final η_{aux}
Temporal / 3D	Pooled frozen history summary	0.01	0.01
Primary images	ImageNet class embedding; CIFAR zero vector	0	0.01
GICO-TF / frozen BézierFlow	CIFAR global context	0	0.01
SANA / SD1.5	Pooled frozen text embedding or global zero vector	0.01	0.05

C.2 OPTIMIZATION AND CHECKPOINT SELECTION

Table 10: Common optimizer settings. Task-specific dropout and score coefficients are in Tab. 9.

Setting	Value
Optimizer	AdamW (Loshchilov & Hutter, 2019); learning rate 10^{-3}, weight decay 10^{-4}, gradient-norm clip 1.
Surrogate / policy updates	2,000 each; up to 64 distinct surrogate groups / 512 policy context–settings per update; microbatches of 8.
Surrogate loss constants	Rank temperature 0.5; exact ties omitted; Huber threshold 1 and regression weight 1.0.
Temperature candidates	$T_{\text{raw}} \in \{0.05, 0.1, 0.5\}$, selected jointly with the surrogate checkpoint.
Auxiliary schedule	Zero for the first 60% of policy updates, then a linear ramp to the final coefficient.
Stochastic training estimates	32 smoothed anchor draws per context for likelihood; 4 fresh policy paths per context for auxiliary refinement.
Selection cadence	Surrogate / deterministic / stochastic: every 20 / 10 / 100 updates, respectively, and the final update.

Surrogate and temperature selection. Hold out 20% of distinct nonuniform densities, rounding up, grouping equivalent aliases, and retaining uniform as the anchor. The collection fixes these identities and the context split. Surrogate fitting and calibration exclude both holdouts. For each measured group, let $U_i = s_{\text{cal}}\widehat{S}_i$ denote utility before scalar normalization and compute

$$w_i(T_{\text{raw}}) = \text{softmax}_i(s_{\text{cal}}\bar{r}_\omega(\rho_i, \xi)/T_{\text{raw}}), \qquad \text{Regret}_\xi(T_{\text{raw}}) = \max_{i\in\mathcal{O}_\xi} U_i - \sum_{i\in\mathcal{O}_\xi} w_i(T_{\text{raw}})U_i,$$

where $\mathcal{O}_\xi$ contains the group's distinct measured references. Selection minimizes the equally weighted mean of context-holdout and density-holdout regrets. The latter uses held-out densities on fitting contexts with their paired uniform anchors. Ties prefer $T_{\text{raw}} = 0.05$, then the earlier checkpoint, then

the smaller temperature. Thus $T_{\mathrm{tilt}} = T_{\mathrm{raw}}/s_{\mathrm{cal}}$ in Eq. (11). The selected surrogate is frozen for both policies.

Policy selection. Both policies use the context holdout and source NFEs only. Let D_j and V_j be the held-out anchor KL and calibrated surrogate utility of checkpoint j. After the training horizon, retain checkpoints with positive auxiliary coefficient and

$$D_j \le (1+\gamma) \min_{k:\, \eta^{(k)}_{\mathrm{aux}} > 0} D_k, \qquad \gamma = 0.15 \text{ (deterministic)}, \quad \gamma = 0.20 \text{ (stochastic)}.$$

Among them, maximize V_j, preferring earlier exact ties. Deterministic selection uses $D_{\mathrm{KL}}(\bar{\boldsymbol{p}}_\xi \| \boldsymbol{p}_{\psi^{(j)}}(\xi))$ and $s_{\mathrm{cal}}\bar{r}_\omega(\rho_{\boldsymbol{p}_{\psi^{(j)}}(\xi)}, \xi)$, reusing one predicted density per group. Utilities and divergences are averaged equally across settings and contexts (classes for ImageNet); auxiliary clipping and standardization are not used for selection.

Stochastic selection uses the full smoothed reference law $\widehat{Q}_{\xi,\epsilon}$ from Appx A.11. With 32 fixed draws per positive-weight reference component, its KL estimate is

$$\widehat{D}^{(j)}_\xi = \sum_i \bar{w}_i \frac{1}{32} \sum_{r=1}^{32} \big[\ell^{(j)}_{ir} + e^{-\ell^{(j)}_{ir}} - 1\big], \qquad \ell^{(j)}_{ir} = \log \widehat{q}_{\xi,\epsilon}(\boldsymbol{z}_{ir}) - \log p_{\psi^{(j)}}(\boldsymbol{z}_{ir} \mid \xi).$$

Both log densities are joint over all 63 coordinates; $\widehat{q}_{\xi,\epsilon}$ is the complete mixture. The estimator is nonnegative samplewise and differs from the coordinate-averaged training NLL. Expected calibrated surrogate utility uses four fixed policy draws per group. Target perturbations and policy draws have separate RNG streams, fixed across checkpoints and independent of training and generation noise.

C.3 Continuous NFE conditioning and exclusion

Conditioning concatenates the frozen native context, one-hot solver identity, and $\log(1+b)$, $\log(1+K_s(b))$. Context statistics use eligible surrogate-fitting rows; budget statistics use distinct fitting solver/NFE settings. Both are standardized once, replacing population SDs below 10^{-6} by one. Global variants zero the standardized context without changing its width or solver/budget inputs. Initial generation noise is never a GICO input.

Held-out NFEs enter neither fitting, normalization, nor checkpoint selection. Deployment changes the continuous budget inputs and decoded node count while keeping networks and normalizers frozen. Tab. 14 lists the NFE splits; COCO transfer is in Appx G.6.

D Datasets and Evaluation

This appendix specifies collection, utility calibration, and evaluation. COCO-specific settings and GICO-TF target preparation are in Appx G.1 and E.5.

D.1 Datasets and collection evidence

For each temporal or molecular backbone, collection samples 256 contexts per NFE, shared across fitting NFEs and split into 204 fitting and 52 held-out contexts. One reference density is assigned to each context: ten contexts per named clock, with six seeded random extras. Each assignment has two generation repeats and a paired uniform anchor; ensemble sizes are in Tab. 11. Molecular trajectories span 16 rollout steps. Repeats and anchors stay in one split, with disjoint physical seeds for ensemble members. The five frozen-field checkpoints give 1,280 context–checkpoint instances; Fig. 4 separately varies the available evidence.

Low-resolution images use official RF++ and EDM-VE checkpoints (Lee et al., 2024; Karras et al., 2022). EDM-VE→RF denotes the frozen EDM field with the RF++-compatible time/velocity adapter. ImageNet-64 uses ILSVRC2012 center crops resized to 64×64 by the EDM preparation pipeline, rather than the box-resized Downsampled ImageNet variant (Chrabaszcz et al., 2017).

Table 11: Available data and collection per NFE. Image totals include both generation repeats.

Dataset	Available units	Collection evidence
Solar-energy-10m	6,648,473 contexts	
Traffic-hourly	14,399,710 contexts	256 contexts; five trajectories per repeat.
Weather-daily	42,402,910 contexts	
Molecule set 1	53,343 contexts	
Molecule set 2	117,285 contexts	256 contexts; 16 trajectories per repeat.
Molecule set 3	54,692 contexts	
CIFAR-10, 32×32	50,000 images	10,000 images including uniform: 400 per density, split 8,000/2,000 over noise/reference panels.
ImageNet-64, 64×64	1,281,167 images	1,000 classes, 40 per density; 64 candidate images per class plus paired uniform; 800/200 class split.

CIFAR-10 uses one zero context, with disjoint generation-noise and real-reference panels for fitting and validation. ImageNet assigns each class to one density; the 40 uniform-assigned classes reuse their outputs as anchors, giving $(1000 + 960) \times 64 = 125{,}440$ images per NFE. Two repeats divide, rather than double, these image allowances. KID uses complete paired feature blocks, with class-matched blocks for ImageNet. Contextual collection counts $2M(2C - C_0)$ trajectories for C contexts, C_0 uniform assignments, and M ensemble members. The collection manifest fixes assignments, splits, seeds, exact grids, and completed-solve counts before fitting.

Table 12: Statistics of the three continuous 3D molecule partitions. Order counts follow triangulene orders $2/3/4/5$; total frames count all XYZ frames; coordinate dimensionality is $3A$ for A atoms.

Dataset	# strata	# formulas	Order counts $2/3/4/5$	Dynamic/ direct	Total frames	# families
Set 1	43	37	13/11/7/12	41/2	81,474	17
Set 2	43	33	7/11/13/12	42/1	176,278	16
Set 3	42	35	12/10/12/8	42/0	83,243	17
All	128	72	32/32/32/32	125/3	340,995	18

Dataset	Atoms mean $\pm$ std.	Atoms median [min, max]	Coord. dim. [min, max]
Set 1	93.8 ± 34.0	86 [44, 156]	[132, 468]
Set 2	101.5 ± 31.4	102 [46, 150]	[138, 450]
Set 3	93.6 ± 30.4	92 [48, 156]	[144, 468]
All	96.3 ± 32.2	96 [44, 156]	[132, 468]

A stratum is a molecular trajectory family. Formula and family totals are deduplicated across partitions; member sets are disjoint. All formulas are hydrocarbons, with $x + y$ atoms for $\mathrm{C}_x\mathrm{H}_y$.

D.2 Empirical utilities and calibration

Tab. 13 defines the paired components and fixed weights. Calibration uses only completed fitting observations outside both context and density holdouts. All required metric components are finite; equivalent reference densities share one identity and must have consistent paired measurements.

Table 13: Paired utility components before scalar calibration.

Study	Higher-is-better paired component v_ℓ	Fixed component weights
Temporal	CRPS (Gneiting & Raftery, 2007) and MASE (Hyndman & Koehler, 2006) log improvements	$0.5, 0.5$
Molecular	Kabsch RMSD (Kabsch, 1976); ensemble and rollout velocity-norm and acceleration-norm W1 log improvements	0.40 for Kabsch; 0.15 for each of four dynamics terms
Primary images; frozen BézierFlow	Uniform-minus-candidate KID (Bińkowski et al., 2018)	1
GICO-TF	Uniform-minus-candidate LPIPS (Zhang et al., 2018) to frozen high-accuracy generated targets	1
COCO	Candidate-minus-uniform IR/VQA, divided by fitting-only component scales (Appx G.2)	$0.5, 0.5$

For temporal/3D error component ℓ, use

$$v_{i\ell} = \log(\widehat{\mu}_{\ell,\mathrm{unif}} + \epsilon_\ell) - \log(\widehat{\mu}_{i\ell} + \epsilon_\ell), \qquad \epsilon_\ell = 10^{-6}\,\mathrm{median}_{\mathrm{calib}:\,\widehat{\mu}_{\ell,\mathrm{unif}}>0}\,\widehat{\mu}_{\ell,\mathrm{unif}}.$$

The floor uses positive uniform measurements from eligible fitting evidence and is then frozen. Error metrics are nonnegative; unbiased KID estimates retain their sign and are never logged. Molecular evaluation U_{comp} uses the four dynamics terms and excludes Kabsch; the fitting reward uses all five components.

For each task/backbone/solver, s_{cal} is the population SD of $\sum_\ell \widetilde{\alpha}_{m,\ell} v_\ell$ over nonuniform fitting candidates, with equal total weight per training NFE and, for ImageNet, per class within NFE. Targets are $\widehat{u}_\ell = v_\ell / s_{\mathrm{cal}}$. This scale stays fixed for validation, test reporting, and transfer; labels are neither mean-centered nor clipped, and uniform utility remains zero. Reference temperature and auxiliary-score normalization are defined separately in Appx C.2 and B.4.

D.3 Evaluation metrics and NFE splits

Temporal/3D full tables use five evaluation seeds and 25 samples per solver. Primary image metrics use matched 50k-sample evaluations over five seeds, with ten IS splits. Appx E.1 defines normalization, aggregation, and reference envelopes. Text-to-image experiments use H100 GPUs; the other studies use A100 GPUs.

KID fitting uses the official EDM Inception detector and its uint8 conversion, with the unbiased cubic-kernel estimator $k(\boldsymbol{f}, \boldsymbol{g}) = (\boldsymbol{f}^\top \boldsymbol{g}/2048 + 1)^3$. Locked primary-image evaluation uses torch-fidelity 0.4.0 (Obukhov et al., 2020), 2,048-dimensional Inception-compatible features, RGB uint8 inputs, bilinear resizing to 299×299, and normalization $(x - 128)/128$. FID (Heusel et al., 2017), IS (Salimans et al., 2016), and precision/recall (Kynkäänniemi et al., 2019) are evaluation-only.

Table 14: Reported NFE splits, in realized generator evaluations. The exclusion concerns GICO fitting and selection; it does not describe pretraining of a frozen backbone or sampler.

Study	Fitting NFEs	Held-out NFEs	Solver / scope
Primary temporal/3D	$4, 8$	$6, 10$	Four solvers, Tab. 6.
Primary low-resolution images	$2, 8$	4	Euler; reported transfer protocol.
Frozen BézierFlow	$4, 8, 10$	6	Frozen two-grid sampler (Appx E.6).
Extended temporal/3D diagnostic	$4, 8, 12, 16$	$6, 10, 14, 20$	Figs. 3, 5, and 7.

E Complete Quantitative Results

E.1 Reporting conventions and reference envelopes

Temporal/3D errors are ratios to matched uniform, which has unit ratios and zero dynamics utility. Each family uses its best unrounded instance per solver/metric, including reversals and coordinate variants. Four-solver aggregates report $\bar{x} = \frac{1}{4}\sum_s x_s$ and $\mathrm{SD} = \sqrt{\frac{1}{4}\sum_s (x_s - \bar{x})^2}$: solver spread, not seed uncertainty. The compact envelope selects a family after solver averaging (Tab. 15); solver-wise family selection would be a stronger oracle. Uniform is omitted from normalized tables but retained in rankings, which use unrounded means.

Primary-image SDs summarize evaluation seeds; their metric-wise reference envelope excludes SharpEuler and stochastic GICO. COCO selects one fixed reference by calibration utility (Appx G.4); optimizer SDs and prompt intervals follow Appx G.1. The separate GICO-TF, frozen BézierFlow, and additional JS tables report point estimates without these uncertainty conventions.

Table 15: Reference families selected after solver aggregation for the compact main tables. Ranks use unrounded means, not a solver-wise oracle.

Dataset	Metric	NFE 4	NFE 6	NFE 8	NFE 10
Solar	rCRPS	OTS	AYS	OTS	OTS
Solar	rMASE	Uniform	AYS	OTS	Late-p
Traffic	rCRPS	OTS	Late-p	Late-p	Late-p
Traffic	rMASE	Late-p	Late-p	GITS	GITS
Weather	rCRPS	Late-p	GITS	GITS	GITS
Weather	rMASE	GITS	GITS	FlowTS	FlowTS
Set 1	rKabsch	GITS	Late-p	Late-p	GITS
Set 1	U_{comp}	OTS	OTS	Uniform	OTS
Set 2	rKabsch	GITS	GITS	GITS	GITS
Set 2	U_{comp}	OTS	OTS	Uniform	AYS
Set 3	rKabsch	GITS	Late-p	FlowTS	FlowTS
Set 3	U_{comp}	OTS	OTS	AYS	AYS

E.2 Temporal sequences

Table 16: Temporal ratios to uniform: final 20k field, mean ± SD over four solvers. Bold/underline mark best/second-best aggregate means. Fitted NFEs: 4/8; transferred NFEs: 6/10.

Metric	Dataset	Clock	NFE 4	NFE 6	NFE 8	NFE 10
			Normalized CRPS ↓ (mean ± solver SD)			
CRPS	Solar	AYS	1.002 ± 0.0042	$\underline{0.993 \pm 0.0113}$	$\underline{0.990 \pm 0.0131}$	0.991 ± 0.0098
		GITS	1.066 ± 0.0210	1.068 ± 0.0194	1.051 ± 0.0159	1.037 ± 0.0166
		OTS	$\underline{0.998 \pm 0.0276}$	0.998 ± 0.0137	$\mathbf{0.987 \pm 0.0164}$	$\mathbf{0.990 \pm 0.0090}$
		Late-p	1.012 ± 0.0458	1.005 ± 0.0228	0.997 ± 0.0198	$\underline{0.991 \pm 0.0104}$
		FlowTS	1.068 ± 0.0239	1.081 ± 0.0244	1.091 ± 0.0382	1.092 ± 0.0435
		GICO	$\mathbf{0.961 \pm 0.0303}$	$\mathbf{0.958 \pm 0.0298}$	1.004 ± 0.0267	1.007 ± 0.0201
CRPS	Traffic	AYS	0.993 ± 0.0045	0.991 ± 0.0058	0.990 ± 0.0048	0.992 ± 0.0040
		GITS	1.053 ± 0.0561	1.030 ± 0.0497	0.993 ± 0.0346	0.975 ± 0.0295
		OTS	$\underline{0.983 \pm 0.0091}$	0.979 ± 0.0096	0.987 ± 0.0083	0.990 ± 0.0091
		Late-p	0.997 ± 0.0205	$\underline{0.971 \pm 0.0113}$	$\underline{0.968 \pm 0.0094}$	$\underline{0.971 \pm 0.0096}$
		FlowTS	1.058 ± 0.0526	1.047 ± 0.0471	1.037 ± 0.0459	1.031 ± 0.0421
		GICO	$\mathbf{0.980 \pm 0.0130}$	$\mathbf{0.963 \pm 0.0124}$	$\mathbf{0.958 \pm 0.0067}$	$\mathbf{0.958 \pm 0.0073}$

Table 16: Continued.

Metric	Dataset	Clock	NFE 4	NFE 6	NFE 8	NFE 10
CRPS	Weather	AYS	0.994 ± 0.0079	0.994 ± 0.0024	0.990 ± 0.0030	0.995 ± 0.0017
		GITS	0.951 ± 0.0568	0.934 ± 0.0550	0.929 ± 0.0378	0.934 ± 0.0232
		OTS	0.980 ± 0.0097	0.982 ± 0.0074	0.995 ± 0.0023	0.991 ± 0.0022
		Late-p	0.950 ± 0.0256	0.960 ± 0.0139	0.967 ± 0.0125	0.974 ± 0.0115
		FlowTS	0.958 ± 0.0525	0.948 ± 0.0546	0.942 ± 0.0498	0.940 ± 0.0496
		GICO	$\mathbf{0.934 \pm 0.0393}$	$\mathbf{0.925 \pm 0.0417}$	$\mathbf{0.914 \pm 0.0428}$	$\mathbf{0.914 \pm 0.0439}$
Normalized MASE ↓ (mean ± solver SD)						
MASE	Solar	AYS	1.001 ± 0.0031	0.997 ± 0.0014	0.995 ± 0.0019	0.993 ± 0.0026
		GITS	1.031 ± 0.0244	1.028 ± 0.0286	1.017 ± 0.0321	1.002 ± 0.0287
		OTS	1.004 ± 0.0049	0.997 ± 0.0021	$\mathbf{0.993 \pm 0.0028}$	0.991 ± 0.0021
		Late-p	1.013 ± 0.0141	1.001 ± 0.0090	0.996 ± 0.0109	$\mathbf{0.986 \pm 0.0075}$
		FlowTS	1.030 ± 0.0265	1.029 ± 0.0301	1.032 ± 0.0402	1.027 ± 0.0402
		GICO	$\mathbf{0.989 \pm 0.0139}$	$\mathbf{0.982 \pm 0.0140}$	1.005 ± 0.0176	1.002 ± 0.0168
MASE	Traffic	AYS	0.986 ± 0.0073	0.985 ± 0.0081	0.985 ± 0.0064	0.987 ± 0.0054
		GITS	0.981 ± 0.0563	0.951 ± 0.0473	0.924 ± 0.0348	0.914 ± 0.0266
		OTS	0.968 ± 0.0097	0.964 ± 0.0115	0.980 ± 0.0101	0.983 ± 0.0129
		Late-p	0.961 ± 0.0315	0.940 ± 0.0135	0.939 ± 0.0122	0.947 ± 0.0145
		FlowTS	0.990 ± 0.0486	0.969 ± 0.0427	0.950 ± 0.0387	0.942 ± 0.0350
		GICO	$\mathbf{0.946 \pm 0.0223}$	$\mathbf{0.928 \pm 0.0264}$	$\mathbf{0.910 \pm 0.0246}$	$\mathbf{0.908 \pm 0.0282}$
MASE	Weather	AYS	0.994 ± 0.0078	0.994 ± 0.0035	0.990 ± 0.0038	0.993 ± 0.0019
		GITS	0.932 ± 0.0504	0.919 ± 0.0510	0.923 ± 0.0356	0.927 ± 0.0213
		OTS	0.982 ± 0.0069	0.982 ± 0.0075	0.994 ± 0.0017	0.990 ± 0.0020
		Late-p	0.953 ± 0.0232	0.959 ± 0.0138	0.965 ± 0.0117	0.970 ± 0.0103
		FlowTS	0.942 ± 0.0481	0.932 ± 0.0505	0.921 ± 0.0455	0.919 ± 0.0466
		GICO	$\mathbf{0.925 \pm 0.0469}$	$\mathbf{0.917 \pm 0.0499}$	$\mathbf{0.901 \pm 0.0482}$	$\mathbf{0.900 \pm 0.0487}$

Table 17: Low-NFE temporal results for Euler at the final 20k checkpoint.

		NFE=4 (seen)		NFE=6 (unseen)		NFE=8 (seen)		NFE=10 (unseen)	
Dataset	Clock	rCRPS↓	rMASE↓	rCRPS↓	rMASE↓	rCRPS↓	rMASE↓	rCRPS↓	rMASE↓
Solar	AYS	1.0072	1.0046	1.0020	**0.9993**	1.0005	0.9980	0.9981	0.9950
	GITS	1.0875	1.0472	1.0869	1.0423	1.0676	1.0305	1.0420	1.0122
	OTS	1.0155	1.0094	1.0090	1.0008	**0.9994**	**0.9971**	**0.9969**	0.9947
	Late-p	1.0447	1.0263	1.0260	1.0098	1.0192	1.0060	1.0056	**0.9931**
	FlowTS	1.0925	1.0477	1.0955	1.0435	1.1082	1.0481	1.1051	1.0439
	GICO	1.0003	1.0120	**0.9962**	1.0060	1.0313	1.0249	1.0297	1.0225
Traffic	AYS	0.9989	0.9894	0.9955	0.9882	0.9940	0.9873	0.9942	0.9891
	GITS	1.0748	0.9849	1.0552	0.9582	1.0038	**0.9242**	0.9805	**0.9138**
	OTS	**0.9978**	0.9792	0.9906	0.9730	0.9962	0.9911	0.9979	0.9948
	Late-p	1.0161	**0.9684**	0.9903	**0.9518**	0.9797	0.9459	0.9773	0.9507
	FlowTS	1.0819	0.9860	1.0680	0.9605	1.0558	0.9393	1.0489	0.9301
	GICO	0.9978	0.9690	**0.9840**	0.9557	**0.9687**	0.9262	**0.9637**	0.9187
Weather	AYS	0.9931	0.9935	0.9966	0.9958	0.9907	0.9898	0.9951	0.9940
	GITS	0.9205	0.9059	**0.9001**	0.8886	0.9052	0.9008	0.9137	0.9094
	OTS	0.9864	0.9868	0.9839	0.9831	0.9966	0.9966	0.9917	0.9905
	Late-p	0.9498	0.9506	0.9556	0.9546	0.9599	0.9576	0.9660	0.9634
	FlowTS	0.9248	**0.8994**	0.9029	**0.8783**	0.8895	**0.8637**	**0.8821**	**0.8568**
	GICO	**0.9187**	0.9100	0.9086	0.9044	**0.8886**	0.8798	0.8886	0.8810

Table 18: Low-NFE temporal results for Heun at the final 20k checkpoint.

		NFE=4 (seen)		NFE=6 (unseen)		NFE=8 (seen)		NFE=10 (unseen)	
Dataset	Clock	rCRPS↓	rMASE↓	rCRPS↓	rMASE↓	rCRPS↓	rMASE↓	rCRPS↓	rMASE↓
Solar	AYS	0.9968	1.0016	0.9739	0.9979	0.9677	0.9960	0.9743	0.9910
	GITS	1.0523	1.0441	1.0703	1.0547	1.0538	1.0525	1.0614	1.0395
	OTS	0.9500	1.0022	0.9743	0.9960	**0.9588**	**0.9934**	**0.9742**	**0.9900**
	Late-p	0.9332	1.0150	0.9667	1.0074	0.9659	1.0036	0.9767	0.9923
	FlowTS	1.0500	1.0438	1.1002	1.0592	1.1288	1.0746	1.1452	1.0696
	GICO	**0.9155**	**0.9872**	**0.9133**	**0.9721**	0.9658	1.0113	0.9837	1.0074
Traffic	AYS	0.9907	0.9912	0.9851	0.9835	0.9856	0.9815	0.9873	0.9834
	GITS	1.1206	1.0709	1.0882	1.0257	1.0446	0.9796	1.0204	0.9555
	OTS	0.9735	0.9746	0.9681	0.9588	0.9738	0.9639	0.9742	0.9617
	Late-p	1.0167	1.0095	0.9672	**0.9472**	**0.9587**	**0.9303**	**0.9587**	**0.9310**
	FlowTS	1.1184	1.0698	1.1065	1.0352	1.0951	1.0099	1.0840	0.9940
	GICO	**0.9679**	**0.9684**	**0.9598**	0.9535	0.9589	0.9405	0.9659	0.9477
Weather	AYS	0.9905	0.9933	0.9951	0.9982	0.9903	0.9936	0.9970	0.9958
	GITS	1.0221	0.9821	0.9963	0.9693	0.9507	0.9461	0.9454	0.9402
	OTS	0.9701	0.9797	0.9800	0.9856	0.9910	0.9917	0.9877	0.9861
	Late-p	**0.9418**	0.9579	0.9624	0.9698	0.9660	0.9694	0.9681	0.9677
	FlowTS	1.0204	0.9814	1.0132	0.9730	1.0002	0.9583	0.9947	0.9530
	GICO	0.9530	**0.9569**	**0.9552**	**0.9622**	**0.9399**	**0.9340**	**0.9393**	**0.9322**

Table 19: Low-NFE temporal results for Midpoint RK2 at the final 20k checkpoint.

		NFE=4 (seen)		NFE=6 (unseen)		NFE=8 (seen)		NFE=10 (unseen)	
Dataset	Clock	rCRPS↓	rMASE↓	rCRPS↓	rMASE↓	rCRPS↓	rMASE↓	rCRPS↓	rMASE↓
Solar	AYS	0.9994	0.9963	0.9989	0.9962	0.9952	0.9931	0.9975	0.9966
	GITS	1.0393	0.9887	1.0361	0.9798	1.0245	**0.9654**	1.0172	**0.9604**
	OTS	1.0114	0.9974	1.0035	0.9973	0.9931	0.9893	0.9941	0.9902
	Late-p	1.0296	0.9896	1.0124	0.9867	0.9945	0.9782	**0.9874**	0.9755
	FlowTS	1.0392	0.9840	1.0391	0.9793	1.0275	0.9660	1.0244	0.9613
	GICO	**0.9594**	**0.9788**	**0.9567**	**0.9738**	**0.9927**	0.9768	0.9902	0.9761
Traffic	AYS	0.9948	0.9913	0.9974	0.9956	0.9963	0.9938	0.9968	0.9949
	GITS	**0.9663**	0.9247	**0.9536**	0.9068	0.9534	0.9047	0.9512	0.9048
	OTS	0.9801	0.9649	0.9862	0.9765	0.9905	0.9846	0.9944	0.9909
	Late-p	0.9686	0.9379	0.9681	0.9443	0.9748	0.9547	0.9823	0.9694
	FlowTS	0.9753	0.9586	0.9778	0.9643	0.9689	0.9474	0.9684	0.9479
	GICO	0.9666	**0.9247**	0.9554	**0.9065**	**0.9510**	**0.8976**	**0.9498**	**0.8957**
Weather	AYS	1.0060	1.0064	0.9939	0.9933	0.9946	0.9938	0.9938	0.9936
	GITS	0.9850	0.9776	0.9775	0.9673	0.9788	0.9671	**0.9656**	**0.9543**
	OTS	0.9929	0.9895	0.9914	0.9884	0.9965	0.9944	0.9931	0.9909
	Late-p	0.9896	0.9849	0.9796	0.9728	0.9879	0.9816	0.9937	0.9880
	FlowTS	0.9971	0.9977	0.9903	0.9916	0.9817	0.9715	0.9827	0.9730
	GICO	**0.9847**	**0.9771**	**0.9707**	**0.9607**	**0.9694**	**0.9566**	0.9703	0.9566

Table 20: Low-NFE temporal results for variable-step AB2 at the final 20k checkpoint.

		NFE=4 (seen)		NFE=6 (unseen)		NFE=8 (seen)		NFE=10 (unseen)	
Dataset	Clock	rCRPS↓	rMASE↓	rCRPS↓	rMASE↓	rCRPS↓	rMASE↓	rCRPS↓	rMASE↓
Solar	AYS	1.0052	1.0029	0.9987	0.9958	0.9969	0.9937	0.9940	0.9904
	GITS	1.0859	1.0434	1.0794	1.0334	1.0561	1.0191	1.0277	0.9943
	OTS	1.0143	1.0087	1.0042	0.9953	**0.9963**	**0.9925**	0.9932	0.9895
	Late-p	1.0405	1.0210	1.0158	0.9996	1.0072	0.9949	**0.9930**	**0.9811**
	FlowTS	1.0909	1.0437	1.0879	1.0343	1.0997	1.0388	1.0947	1.0331
	GICO	**0.9691**	**0.9772**	**0.9673**	**0.9751**	1.0267	1.0073	1.0236	1.0039
Traffic	AYS	0.9870	0.9738	0.9846	0.9734	0.9858	0.9766	0.9880	0.9810
	GITS	1.0507	0.9433	1.0246	0.9140	0.9718	0.8870	**0.9462**	0.8820
	OTS	**0.9792**	0.9540	0.9711	0.9476	0.9890	0.9809	0.9917	0.9856
	Late-p	0.9857	0.9289	0.9602	0.9173	0.9588	0.9239	0.9639	0.9383
	FlowTS	1.0579	0.9442	1.0373	0.9159	1.0294	0.9023	1.0244	0.8970
	GICO	0.9857	**0.9235**	**0.9527**	**0.8979**	**0.9542**	**0.8772**	0.9506	**0.8713**
Weather	AYS	0.9844	0.9845	0.9901	0.9887	0.9861	0.9844	0.9924	0.9904
	GITS	**0.8751**	0.8622	**0.8620**	0.8512	0.8821	0.8772	0.9094	0.9028
	OTS	0.9716	0.9717	0.9709	0.9690	0.9948	0.9943	0.9928	0.9905
	Late-p	0.9185	0.9198	0.9408	0.9382	0.9554	0.9508	0.9672	0.9626
	FlowTS	0.8908	0.8891	0.8858	0.8862	0.8951	0.8887	0.8989	0.8918
	GICO	0.8792	**0.8555**	0.8643	**0.8404**	**0.8596**	**0.8326**	**0.8569**	**0.8300**

E.3 Molecular coordinates

Table 21: 3D results: final 20k field, mean ± SD over four solvers. U_{comp} excludes Kabsch; bold/underline mark best/second-best. Fitted NFEs: 4/8; transferred NFEs: 6/10.

Metric	Dataset	Clock	NFE 4	NFE 6	NFE 8	NFE 10
Normalized Kabsch RMSD ↓ (mean ± solver SD)						
Kabsch	Set 1	AYS	0.998 ± 0.0044	0.997 ± 0.0038	0.998 ± 0.0018	0.998 ± 0.0020
		GITS	0.991 ± 0.0999	0.991 ± 0.1181	0.994 ± 0.0961	0.991 ± 0.0792
		OTS	0.994 ± 0.0091	0.993 ± 0.0071	0.999 ± 0.0018	0.997 ± 0.0019
		Late-p	1.000 ± 0.0446	0.989 ± 0.0175	0.989 ± 0.0111	0.991 ± 0.0075
		FlowTS	1.002 ± 0.0905	1.011 ± 0.1054	0.995 ± 0.1233	0.996 ± 0.1230
		GICO	$\mathbf{0.945 \pm 0.0408}$	$\mathbf{0.935 \pm 0.0505}$	$\mathbf{0.926 \pm 0.0575}$	$\mathbf{0.923 \pm 0.0602}$
Kabsch	Set 2	AYS	0.996 ± 0.0044	0.997 ± 0.0040	0.997 ± 0.0027	0.998 ± 0.0020
		GITS	0.984 ± 0.0927	0.983 ± 0.1078	0.986 ± 0.0876	0.985 ± 0.0727
		OTS	0.992 ± 0.0081	0.992 ± 0.0076	0.997 ± 0.0025	0.997 ± 0.0011
		Late-p	0.996 ± 0.0430	0.986 ± 0.0163	0.987 ± 0.0105	0.988 ± 0.0077
		FlowTS	1.007 ± 0.0753	1.012 ± 0.0880	0.998 ± 0.1009	1.000 ± 0.1002
		GICO	$\mathbf{0.946 \pm 0.0423}$	$\mathbf{0.937 \pm 0.0511}$	$\mathbf{0.929 \pm 0.0568}$	$\mathbf{0.927 \pm 0.0605}$
Kabsch	Set 3	AYS	0.997 ± 0.0047	0.997 ± 0.0043	0.997 ± 0.0028	0.997 ± 0.0030
		GITS	0.991 ± 0.0990	0.990 ± 0.1171	0.987 ± 0.0921	0.984 ± 0.0748
		OTS	0.991 ± 0.0095	0.991 ± 0.0088	0.996 ± 0.0026	0.996 ± 0.0032
		Late-p	0.998 ± 0.0450	0.986 ± 0.0178	0.983 ± 0.0090	0.987 ± 0.0066
		FlowTS	1.001 ± 0.0911	1.009 ± 0.1041	0.983 ± 0.1234	0.981 ± 0.1236
		GICO	$\mathbf{0.946 \pm 0.0413}$	$\mathbf{0.936 \pm 0.0512}$	$\mathbf{0.923 \pm 0.0521}$	$\mathbf{0.921 \pm 0.0552}$
Dynamics utility U_{comp} ↑ (mean ± solver SD)						
U_{comp}	Set 1	AYS	-0.001 ± 0.0106	$+0.001 \pm 0.0097$	-0.002 ± 0.0050	-0.001 ± 0.0034
		GITS	-0.035 ± 0.1745	-0.045 ± 0.2024	-0.050 ± 0.1578	-0.037 ± 0.1259
		OTS	$+0.002 \pm 0.0227$	$+0.001 \pm 0.0151$	-0.002 ± 0.0038	$+0.000 \pm 0.0037$
		Late-p	-0.030 ± 0.0520	-0.008 ± 0.0128	-0.004 ± 0.0067	-0.003 ± 0.0035
		FlowTS	-0.048 ± 0.1642	-0.067 ± 0.1970	-0.051 ± 0.2246	-0.052 ± 0.2252
		GICO	$\mathbf{+0.064 \pm 0.0466}$	$\mathbf{+0.078 \pm 0.0574}$	$\mathbf{+0.089 \pm 0.0661}$	$\mathbf{+0.093 \pm 0.0690}$

Table 21: Continued.

Metric	Dataset	Clock	NFE 4	NFE 6	NFE 8	NFE 10
U_{comp}	Set 2	AYS	$+0.001 \pm 0.0101$	$+0.001 \pm 0.0091$	-0.000 ± 0.0057	$+0.001 \pm 0.0032$
		GITS	-0.033 ± 0.1661	-0.042 ± 0.1934	-0.046 ± 0.1480	-0.033 ± 0.1200
		OTS	$+0.003 \pm 0.0208$	$+0.003 \pm 0.0145$	-0.001 ± 0.0040	$+0.000 \pm 0.0028$
		Late-p	-0.027 ± 0.0506	-0.006 ± 0.0126	-0.003 ± 0.0062	-0.001 ± 0.0040
		FlowTS	-0.054 ± 0.1511	-0.067 ± 0.1841	-0.059 ± 0.2047	-0.061 ± 0.2056
		GICO	$\mathbf{+0.057 \pm 0.0435}$	$\mathbf{+0.069 \pm 0.0557}$	$\mathbf{+0.079 \pm 0.0619}$	$\mathbf{+0.082 \pm 0.0666}$
U_{comp}	Set 3	AYS	-0.000 ± 0.0113	$+0.000 \pm 0.0099$	$+0.000 \pm 0.0056$	$+0.001 \pm 0.0036$
		GITS	-0.043 ± 0.1848	-0.054 ± 0.2166	-0.058 ± 0.1672	-0.043 ± 0.1347
		OTS	$+0.003 \pm 0.0245$	$+0.001 \pm 0.0155$	-0.001 ± 0.0039	$+0.000 \pm 0.0034$
		Late-p	-0.034 ± 0.0575	-0.008 ± 0.0126	-0.003 ± 0.0069	-0.002 ± 0.0046
		FlowTS	-0.052 ± 0.1777	-0.072 ± 0.2133	-0.057 ± 0.2400	-0.056 ± 0.2434
		GICO	$\mathbf{+0.067 \pm 0.0446}$	$\mathbf{+0.079 \pm 0.0568}$	$\mathbf{+0.087 \pm 0.0600}$	$\mathbf{+0.090 \pm 0.0643}$

Table 22: Low-NFE continuous 3D results for Euler at the final 20k checkpoint.

		NFE=4 (seen)		NFE=6 (unseen)		NFE=8 (seen)		NFE=10 (unseen)	
Dataset	Clock	rKabsch↓	U_{comp} ↑	rKabsch↓	U_{comp} ↑	rKabsch↓	U_{comp} ↑	rKabsch↓	U_{comp} ↑
Set 1	AYS	1.0022	−0.01063	1.0008	−0.00728	1.0004	−0.00686	0.9984	−0.00321
	GITS	0.9345	+0.06831	0.9185	+0.07924	0.9311	+0.04632	0.9318	+0.04750
	OTS	0.9988	−0.01121	0.9986	−0.00992	0.9994	−0.00543	0.9971	−0.00196
	Late-p	0.9769	−0.00042	0.9803	−0.00231	0.9817	−0.00363	0.9854	−0.00374
	FlowTS	0.9436	+0.05839	0.9327	+0.07300	0.9051	+0.09644	0.8985	+0.10016
	GICO	**0.9297**	**+0.08015**	**0.9065**	**+0.10971**	**0.8898**	**+0.12894**	**0.8815**	**+0.13902**
Set 2	AYS	1.0004	−0.00893	1.0004	−0.00694	0.9991	−0.00503	0.9982	−0.00205
	GITS	0.9345	+0.06200	0.9184	+0.07390	0.9303	+0.04589	0.9317	+0.04837
	OTS	0.9969	−0.01022	0.9974	−0.00831	0.9985	−0.00374	0.9976	−0.00271
	Late-p	0.9754	−0.00079	0.9782	−0.00153	0.9820	−0.00207	0.9839	−0.00109
	FlowTS	0.9567	+0.04073	0.9452	+0.05991	0.9248	+0.08073	0.9245	+0.08057
	GICO	**0.9305**	**+0.07061**	**0.9079**	**+0.09859**	**0.8893**	**+0.12120**	**0.8810**	**+0.13079**
Set 3	AYS	1.0007	−0.01030	1.0004	−0.00951	0.9987	−0.00512	0.9977	−0.00162
	GITS	0.9379	+0.06505	0.9208	+0.07614	0.9331	+0.04247	0.9328	+0.04581
	OTS	0.9970	−0.01361	0.9967	−0.01189	0.9990	−0.00459	0.9978	−0.00183
	Late-p	0.9774	−0.00432	0.9780	−0.00402	0.9803	−0.00270	0.9838	−0.00264
	FlowTS	0.9529	+0.05297	0.9381	+0.07149	0.9126	+0.08294	0.9055	+0.09057
	GICO	**0.9336**	**+0.08227**	**0.9094**	**+0.11080**	**0.8962**	**+0.12066**	**0.8882**	**+0.12982**

Table 23: Low-NFE continuous 3D results for Heun at the final 20k checkpoint.

		NFE=4 (seen)		NFE=6 (unseen)		NFE=8 (seen)		NFE=10 (unseen)	
Dataset	Clock	rKabsch↓	U_{comp} ↑	rKabsch↓	U_{comp} ↑	rKabsch↓	U_{comp} ↑	rKabsch↓	U_{comp} ↑
Set 1	AYS	0.9920	+0.01698	0.9924	+0.01733	**0.9969**	**+0.00661**	0.9975	+0.00478
	GITS	1.1542	−0.32906	1.1864	−0.38894	1.1535	−0.32088	1.1229	−0.25324
	OTS	**0.9838**	**+0.04008**	**0.9851**	**+0.02702**	0.9975	+0.00393	**0.9962**	**+0.00603**
	Late-p	1.0707	−0.11688	1.0112	−0.02615	1.0011	−0.01284	0.9988	−0.00593
	FlowTS	1.1548	−0.32960	1.1919	−0.40612	1.2031	−0.43541	1.2033	−0.43756
	GICO	0.9888	+0.01185	0.9910	+0.01310	1.0025	+0.00311	1.0020	+0.00536

Table 23: Continued.

		NFE=4 (seen)		NFE=6 (unseen)		NFE=8 (seen)		NFE=10 (unseen)	
Dataset	Clock	rKabsch↓	U_{comp} ↑	rKabsch↓	U_{comp} ↑	rKabsch↓	U_{comp} ↑	rKabsch↓	U_{comp} ↑
Set 2	AYS	0.9916	+0.01751	0.9911	+0.01678	0.9943	**+0.00917**	0.9972	**+0.00559**
	GITS	1.1355	−0.31412	1.1603	−0.37110	1.1312	−0.30016	1.1049	−0.23869
	OTS	**0.9852**	**+0.03783**	**0.9827**	**+0.02702**	0.9950	+0.00549	0.9951	+0.00470
	Late-p	1.0644	−0.11217	1.0056	−0.02325	0.9964	−0.00922	0.9937	−0.00422
	FlowTS	1.1360	−0.31468	1.1637	−0.38501	1.1704	−0.41150	1.1715	−0.41552
	GICO	0.9955	+0.01049	0.9946	+0.01151	**0.9892**	+0.00294	**0.9893**	+0.00291
Set 3	AYS	0.9915	+0.01899	0.9909	+0.01684	0.9937	**+0.00940**	0.9945	+0.00609
	GITS	1.1518	−0.35523	1.1830	−0.42367	1.1403	−0.34558	1.1089	−0.27442
	OTS	0.9818	**+0.04441**	**0.9815**	**+0.02737**	0.9927	+0.00536	0.9913	+0.00514
	Late-p	1.0691	−0.13134	1.0063	−0.02687	0.9920	−0.01319	0.9894	−0.00698
	FlowTS	1.1524	−0.35594	1.1869	−0.43977	1.1887	−0.46830	1.1879	−0.47199
	GICO	**0.9814**	+0.02421	0.9826	+0.02306	**0.9827**	+0.00924	**0.9835**	**+0.00774**

Table 24: Low-NFE continuous 3D results for Midpoint RK2 at the final 20k checkpoint.

		NFE=4 (seen)		NFE=6 (unseen)		NFE=8 (seen)		NFE=10 (unseen)	
Dataset	Clock	rKabsch↓	U_{comp} ↑	rKabsch↓	U_{comp} ↑	rKabsch↓	U_{comp} ↑	rKabsch↓	U_{comp} ↑
Set 1	AYS	1.0020	−0.00532	1.0011	−0.00427	1.0002	−0.00314	1.0012	−0.00371
	GITS	0.9842	+0.00589	0.9784	+0.00981	0.9811	+0.00664	0.9819	+0.00549
	OTS	1.0057	−0.01853	1.0019	−0.00931	1.0011	−0.00535	1.0006	−0.00386
	Late-p	1.0001	−0.01930	0.9994	−0.01165	0.9982	−0.00679	0.9991	−0.00466
	FlowTS	0.9833	+0.00742	**0.9758**	+0.01581	0.9715	+0.02434	0.9712	+0.02475
	GICO	**0.9762**	**+0.03165**	0.9761	**+0.03365**	**0.9576**	**+0.04967**	**0.9593**	**+0.04858**
Set 2	AYS	1.0012	−0.00370	1.0002	−0.00294	1.0001	−0.00437	1.0005	−0.00240
	GITS	0.9783	+0.01074	0.9722	+0.01321	0.9765	+0.00756	0.9793	+0.00519
	OTS	1.0033	−0.01513	1.0015	−0.00848	1.0009	−0.00494	0.9977	−0.00117
	Late-p	0.9949	−0.01435	0.9959	−0.00931	0.9968	−0.00637	0.9977	−0.00339
	FlowTS	0.9788	+0.01011	**0.9701**	+0.01808	**0.9681**	+0.02182	**0.9670**	+0.02411
	GICO	**0.9739**	**+0.02543**	0.9758	**+0.02101**	0.9798	**+0.03544**	0.9837	**+0.03049**
Set 3	AYS	1.0019	−0.00533	1.0012	−0.00323	0.9998	−0.00292	1.0014	−0.00326
	GITS	0.9855	+0.00331	**0.9781**	+0.01225	0.9711	+0.00885	0.9753	+0.00505
	OTS	1.0040	−0.01707	1.0031	−0.00840	0.9986	−0.00381	0.9999	−0.00392
	Late-p	1.0000	−0.01923	0.9987	−0.01002	0.9908	−0.00253	0.9959	−0.00448
	FlowTS	**0.9844**	+0.00511	0.9787	+0.01428	0.9613	+0.02724	**0.9620**	+0.02626
	GICO	0.9864	**+0.02814**	0.9874	**+0.02524**	**0.9612**	**+0.05214**	0.9634	**+0.05095**

Table 25: Low-NFE continuous 3D results for variable-step AB2 at the final 20k checkpoint.

		NFE=4 (seen)		NFE=6 (unseen)		NFE=8 (seen)		NFE=10 (unseen)	
Dataset	Clock	rKabsch↓	U_{comp} ↑	rKabsch↓	U_{comp} ↑	rKabsch↓	U_{comp} ↑	rKabsch↓	U_{comp} ↑
Set 1	AYS	0.9949	−0.00519	0.9945	−0.00267	0.9964	−0.00269	0.9958	+0.00001
	GITS	0.8909	+0.11678	0.8805	+0.11991	0.9083	+0.06716	0.9269	+0.05090
	OTS	0.9859	−0.00216	0.9876	−0.00255	0.9965	−0.00135	0.9959	+0.00063
	Late-p	0.9508	+0.01782	0.9657	+0.00875	0.9745	+0.00558	0.9826	+0.00302
	FlowTS	0.9270	+0.07074	0.9448	+0.05013	0.9007	+0.11235	0.9099	+0.10634
	GICO	**0.8857**	**+0.13216**	**0.8675**	**+0.15542**	**0.8548**	**+0.17250**	**0.8509**	**+0.17833**

Table 25: Continued.

		NFE=4 (seen)		NFE=6 (unseen)		NFE=8 (seen)		NFE=10 (unseen)	
Dataset	Clock	rKabsch↓	U_{comp} ↑	rKabsch↓	U_{comp} ↑	rKabsch↓	U_{comp} ↑	rKabsch↓	U_{comp} ↑
Set 2	AYS	0.9927	−0.00261	0.9944	−0.00147	0.9942	−0.00137	0.9949	+0.00107
	GITS	0.8896	+0.10977	0.8794	+0.11549	0.9057	+0.06221	0.9228	+0.05169
	OTS	0.9838	−0.00029	0.9867	+0.00006	0.9950	−0.00103	0.9962	−0.00064
	Late-p	0.9483	+0.01908	0.9631	+0.01157	0.9718	+0.00714	0.9782	+0.00590
	FlowTS	0.9551	+0.04716	0.9703	+0.03855	0.9290	+0.07259	0.9371	+0.06621
	GICO	**0.8854**	**+0.12189**	**0.8680**	**+0.14612**	**0.8575**	**+0.15569**	**0.8533**	**+0.16241**
Set 3	AYS	0.9924	−0.00355	0.9939	−0.00266	0.9938	−0.00105	0.9939	+0.00246
	GITS	0.8887	+0.11614	0.8774	+0.11816	0.9017	+0.06270	0.9199	+0.05115
	OTS	0.9824	−0.00074	0.9845	−0.00179	0.9954	−0.00066	0.9953	+0.00117
	Late-p	0.9467	+0.01741	0.9609	+0.00825	0.9697	+0.00638	0.9780	+0.00528
	FlowTS	0.9134	+0.08822	0.9331	+0.06480	0.8684	+0.12907	0.8702	+0.13270
	GICO	**0.8845**	**+0.13312**	**0.8655**	**+0.15553**	**0.8514**	**+0.16459**	**0.8475**	**+0.17154**

E.4 Primary low-resolution image results

Tab. 26 expands the primary KID-supervised image comparison in Sec. 6.1. SharpEuler and stochastic GICO are shown separately from the fixed-reference envelope.

Table 26: Image metrics (50k samples, Euler, five seeds): reference/SharpEuler and deterministic/stochastic GICO. Bold/underline mark best/second-best; ties share ranks.

NFE	Clock	FID50k↓	IS mean↑	Precision↑	Recall↑
CIFAR-10 RF++					
2 (seen)	Best reference	2.466	9.639	0.677	0.602
	SharpEuler	2.438	9.584	0.677	0.602
	GICO (deterministic)	**2.402**	**9.888**	0.675	**0.606**
	GICO (stochastic)	2.478	9.594	**0.680**	0.601
4 (unseen)	Best reference	2.290	9.667	**0.679**	0.603
	SharpEuler	2.225	**9.688**	0.676	**0.607**
	GICO (deterministic)	**2.218**	9.670	0.674	0.606
	GICO (stochastic)	2.266	9.676	0.678	0.606
8 (seen)	Best reference	2.215	9.711	0.675	0.609
	SharpEuler	2.187	9.748	**0.678**	0.608
	GICO (deterministic)	2.210	**9.909**	0.674	**0.619**
	GICO (stochastic)	**2.002**	9.736	**0.678**	0.607
CIFAR-10 EDM-VE→RF					
2 (seen)	Best reference	128.943	3.368	**0.365**	0.121
	SharpEuler	124.802	3.819	0.322	**0.128**
	GICO (deterministic)	**121.247**	**3.837**	0.337	**0.128**
	GICO (stochastic)	130.765	3.584	0.311	0.124
4 (unseen)	Best reference	41.955	6.883	0.492	0.324
	SharpEuler	39.913	**6.984**	**0.501**	0.324
	GICO (deterministic)	**39.542**	6.950	0.500	**0.328**
	GICO (stochastic)	49.015	6.473	0.471	0.308
8 (seen)	Best reference	15.450	8.323	0.593	0.480
	SharpEuler	14.568	8.355	0.602	**0.490**
	GICO (deterministic)	**14.244**	**8.359**	**0.606**	0.489
	GICO (stochastic)	18.303	8.161	0.581	0.478
ImageNet-64 RF++					
2 (seen)	Best reference	3.655	42.905	0.665	0.630
	SharpEuler	4.531	42.382	0.664	0.628
	GICO (deterministic)	**3.334**	**43.623**	0.663	**0.631**
	GICO (stochastic)	3.661	43.233	**0.668**	0.625

Table 26: Continued.

NFE	Clock	FID50k↓	IS mean↑	Precision↑	Recall↑
4 (unseen)	Best reference	3.324	43.233	0.668	**0.633**
	SharpEuler	4.187	42.381	0.671	0.631
	GICO (deterministic)	3.061	**44.047**	**0.678**	0.631
	GICO (stochastic)	**2.986**	44.024	0.670	0.631
8 (seen)	Best reference	3.200	43.731	0.671	0.634
	SharpEuler	4.070	42.333	0.673	0.631
	GICO (deterministic)	**2.930**	**45.881**	**0.690**	**0.642**
	GICO (stochastic)	3.192	44.199	0.670	0.631
ImageNet-64 EDM-VE→RF					
2 (seen)	Best reference	174.098	4.299	**0.257**	0.068
	SharpEuler	186.759	3.620	0.129	0.036
	GICO (deterministic)	**152.510**	**5.794**	0.255	**0.099**
	GICO (stochastic)	172.350	4.060	0.202	0.044
4 (unseen)	Best reference	59.235	15.191	0.246	0.341
	SharpEuler	89.421	9.967	0.262	0.246
	GICO (deterministic)	**43.718**	**17.326**	**0.280**	0.353
	GICO (stochastic)	74.732	11.956	0.262	**0.373**
8 (seen)	Best reference	21.820	28.146	0.452	0.497
	SharpEuler	23.047	28.160	0.481	0.523
	GICO (deterministic)	**15.077**	**32.449**	**0.531**	**0.539**
	GICO (stochastic)	21.507	24.545	0.430	0.519

E.5 Additional teacher-fidelity comparison

GICO-TF uses uniform-minus-candidate VGG-LPIPS to high-accuracy outputs of the same frozen generator with matched noise/class pairs. Targets and feature models are fixed across clocks; target pairs and noise panels are split-disjoint. LPIPS is measured on decoded float32 tensors without resizing, clamping, or uint8 conversion. Tab. 27 evaluates FID_{50k} using the shared GICO fitting procedure.

Table 27: Additional CIFAR-10 teacher-fidelity comparison: FID_{50k} (lower is better). GICO-TF uses LPIPS supervision and fixes the original field and Euler (RK1) solver; BézierFlow is a broader trajectory-optimization reference.

NFE	Base RK1	LD3 adaptation	BézierFlow	GICO-TF
4	55.50	53.79	**23.20**	48.574
6	28.23	32.24	**14.36**	23.153
8	19.46	28.71	**10.53**	16.068
10	15.30	22.24	**8.49**	12.324

GICO-TF improves FID over native RK1 and the LD3 adaptation at every evaluated NFE. BézierFlow, which also changes the trajectory transformation, achieves the lowest absolute FID.

E.6 KID-supervised clock optimization on frozen BézierFlow

This experiment fits KID-supervised GICO at NFEs 4/8/10 and evaluates transfer to NFE 6. The holdout applies to GICO, not BézierFlow pretraining. Its independently evaluated baselines are kept separate from the LPIPS comparison in Appx E.5.

The Reflow1 EMA field, BézierFlow transformation, and its two learned grids are frozen. GICO uses global context and warps both grids by interpolation over normalized node index at its inverse-CDF clock nodes. Uniform recovers the original two-grid sampler. Endpoints, Euler update rule, and realized field-evaluation count stay fixed; only the placement along the frozen grids changes.

Table 28: Clock optimization on frozen BézierFlow. GICO uses KID supervision at NFEs 4/8/10 and transfers to the GICO-held-out NFE 6. Relative change is $100(\mathrm{FID}_{+\mathrm{GICO}}/\mathrm{FID}_{\mathrm{frozen}} - 1)$; negative values indicate improvement. Bold marks the lower FID within each row.

NFE	GICO budget status	Frozen BézierFlow	Frozen BézierFlow + GICO	Relative change
4	Fitted	20.92	**16.65**	−20.4%
6	Transfer	12.10	**11.01**	−9.0%
8	Fitted	8.33	**7.45**	−10.6%
10	Fitted	**6.26**	7.26	+16.0%

FID decreases by 9.0–20.4% at NFEs 4/6/8, including the held-out NFE, but increases by 16.0% at NFE 10. Composition therefore gives budget-dependent gains rather than a uniform improvement over the frozen sampler.

F Diagnostics and Examples

The following diagnostics use the NFE splits in Tab. 14.

F.1 Robustness across vector-field maturity and NFE

Fig. 5 compares frozen temporal/3D checkpoints from 4k to 20k steps using the extended NFE split in Tab. 14. Circles and diamonds mark seen and held-out NFEs; bands show the 10th–90th percentiles over matched dataset–solver cells. Aggregate effects remain favorable but vary non-monotonically with checkpoint maturity and NFE.

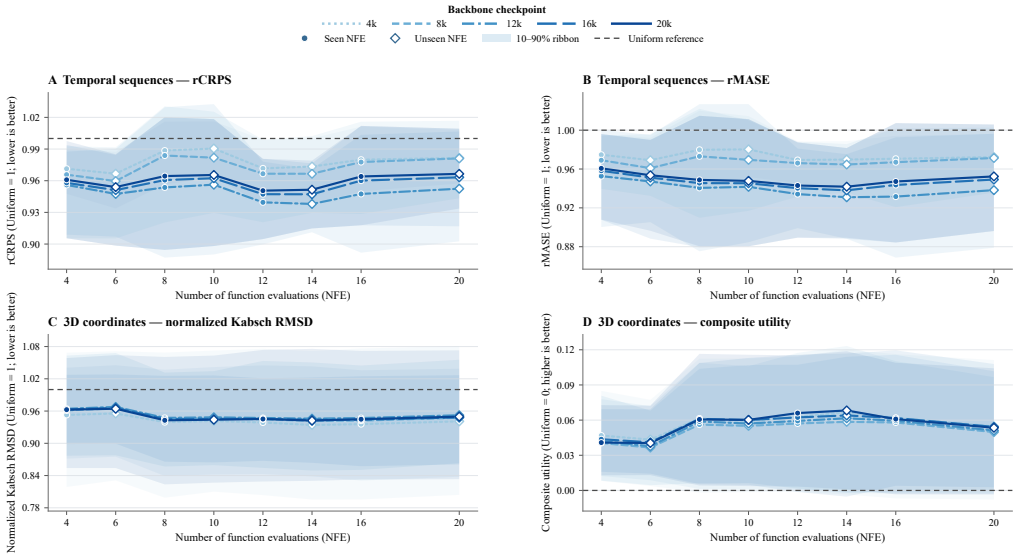

Figure 5: Robustness to vector-field maturity and NFE. Curves aggregate matched temporal and continuous 3D cells; the uniform clock is the dashed reference.

F.2 Density representation and context conditioning

We ablate density/inverse-CDF placement and context conditioning. **Direct-grid** predicts padded macro-step nodes, regularizes monotonicity, and projects to pinned, strictly ordered endpoints. **Context-free** removes temporal/molecular histories and ImageNet classes but retains the shared solver/NFE inputs; CIFAR-10 is already context-free. All variants share supervision, optimization budgets, selection, and locked cells. Ratios to GICO average temporal CRPS/MASE, 3D Kabsch RMSD, and image FID50k.

Table 29: Density and context ablations. Ratios to GICO; larger is worse.

Ablation mode	Context conditioning	Density representation	Relative metric ratio ↓
GICO	✓	✓	**1.000**
Direct-grid	✓	✘	1.300
Context-free	✘	✓	1.187
Direct-grid + context-free	✘	✘	1.364

F.3 Deterministic and stochastic causal image policies

Tab. 30 compares mean densities and stochastic diversity; most ImageNet variation occurs within classes. Fig. 6 also shows SharpEuler. Mean-density proximity does not imply identical numerical outputs (Appx A.11).

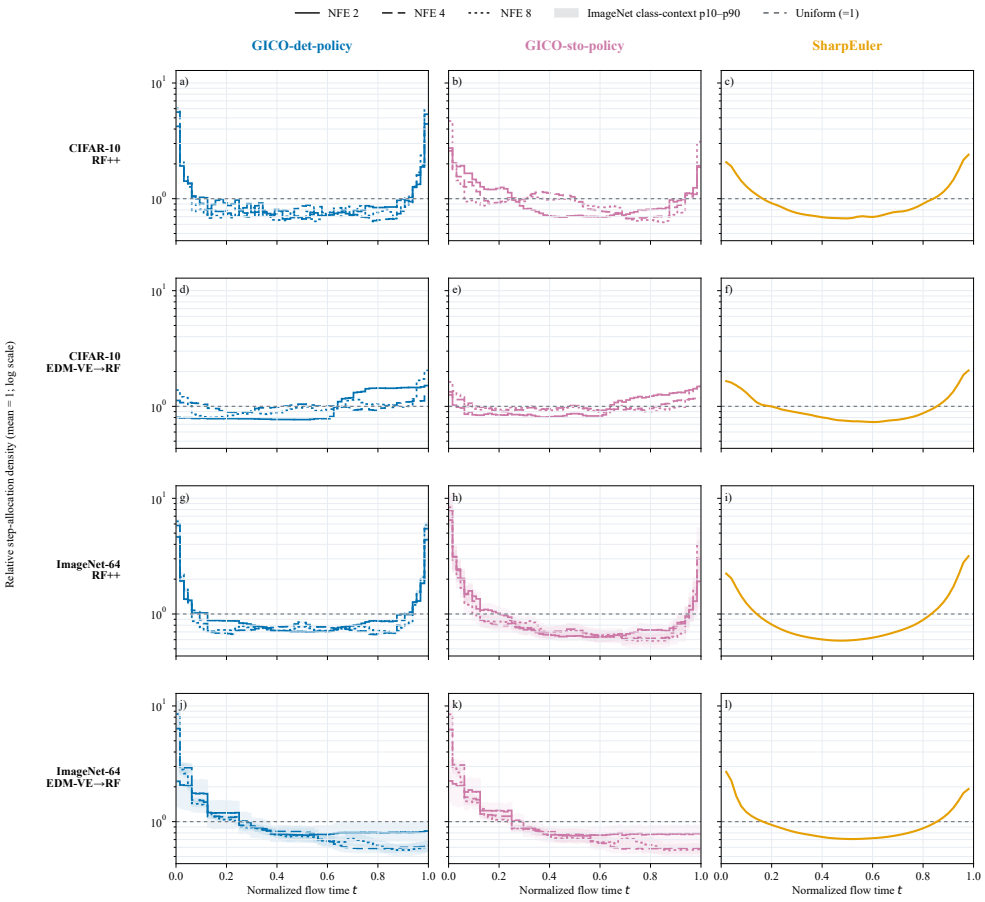

Figure 6: Relative step-allocation densities for deterministic GICO, stochastic GICO, and SharpEuler (mean one; log scale). Line style denotes NFE; ImageNet bands show class-context p10–p90 range.

Table 30: Mean-density proximity and stochastic diversity

Dataset / backbone	NFE	TV	Unique	Class median	Mode share	Class share	Block share
CIFAR-10 / RF++	2	0.0974539	49,999	–	0.004%	–	8.21687×10^{-4}
	4	0.139138	50,000	–	0.002%	–	9.24657×10^{-4}
	8	0.125722	50,000	–	0.002%	–	1.01372×10^{-3}
CIFAR-10 / EDM-VE→RF	2	0.0505711	50,000	–	0.002%	–	9.77336×10^{-4}
	4	0.0183775	50,000	–	0.002%	–	1.07867×10^{-3}
	8	0.0500309	49,999	–	0.004%	–	1.12816×10^{-3}
ImageNet-64 / RF++	2	0.122496	50,000	50	0.002%	0.0227742	7.52392×10^{-4}
	4	0.114418	50,000	50	0.002%	0.0232749	1.08367×10^{-3}
	8	0.0990331	50,000	50	0.002%	0.0238508	9.69067×10^{-4}
ImageNet-64 / EDM-VE→RF	2	0.013232	50,000	50	0.002%	0.160475	8.53779×10^{-4}
	4	0.0158727	50,000	50	0.002%	0.0761643	9.28779×10^{-4}
	8	0.0112978	50,000	50	0.002%	0.0611737	9.00911×10^{-4}

TV (total variation) compares cell-mean masses. Unique counts describe 50,000 stochastic mean-one density rows per cell, using nearest-even quantization at 10^{-6}; all counts are unchanged at 10^{-5} and 10^{-7}. ImageNet class medians are out of 50 draws per class. Class and block shares are separate marginal fractions of variance and are not additive. These statistics describe clock diversity, not image quality.

F.4 Policy latency

Tabs. 31–33 measure policy overhead separately from field latency (Tab. 34). "Policy" measures the predictor; "Complete" also includes input and context embedding processing and density-to-grid conversion. Stochastic timings include the 63-coordinate causal decoder.

Table 31: Temporal/3D deterministic policy latency on NVIDIA A100.

Temporal dataset / 3D set	NFE	Batch 1 (ms) Policy	Batch 1 (ms) Complete	Batch 64 (ms) Policy	Batch 64 (ms) Complete	Batch 64 samples/s
Temporal sequences						
Solar	4	0.204	2.301	0.205	2.360	27,115
Solar	8	0.203	2.288	0.209	2.357	27,150
Solar	12	0.204	2.299	0.209	2.337	27,388
Solar	16	0.203	2.327	0.208	2.355	27,174
Traffic	4	0.202	2.320	0.207	2.348	27,257
Traffic	8	0.202	2.331	0.206	2.361	27,109
Traffic	12	0.205	2.306	0.206	2.375	26,951
Traffic	16	0.206	2.285	0.207	2.342	27,328
Weather	4	0.204	2.280	0.206	2.347	27,269
Weather	8	0.205	2.310	0.209	2.332	27,448
Weather	12	0.204	2.312	0.208	2.349	27,245
Weather	16	0.204	2.330	0.207	2.358	27,139
Continuous 3D coordinates						
Set 1	4	0.201	2.323	0.206	2.348	27,257
Set 1	8	0.203	2.309	0.206	2.354	27,186
Set 1	12	0.204	2.294	0.207	2.356	27,162
Set 1	16	0.205	2.300	0.206	2.335	27,412
Set 2	4	0.204	2.298	0.207	2.347	27,269
Set 2	8	0.203	2.297	0.207	2.345	27,293
Set 2	12	0.204	2.306	0.207	2.345	27,293
Set 2	16	0.204	2.314	0.207	2.350	27,233
Set 3	4	0.201	2.331	0.207	2.333	27,436
Set 3	8	0.203	2.301	0.206	2.367	27,033
Set 3	12	0.202	2.302	0.206	2.355	27,174
Set 3	16	0.204	2.300	0.205	2.344	27,305

Table 32: Image deterministic policy latency on NVIDIA A100.

		Batch 1 latency (ms)		Batch 64 latency (ms)		Batch 64
Dataset / backbone	NFE	Policy	Complete	Policy	Complete	samples/s
CIFAR-10 RF++	2	0.231	0.451	0.231	0.454	141,084
CIFAR-10 RF++	4	0.232	0.457	0.230	0.468	136,761
CIFAR-10 RF++	8	0.234	0.457	0.234	0.466	137,363
CIFAR-10 EDM-VE→RF	2	0.230	0.445	0.232	0.454	141,084
CIFAR-10 EDM-VE→RF	4	0.230	0.463	0.233	0.466	137,363
CIFAR-10 EDM-VE→RF	8	0.240	0.449	0.236	0.466	137,363
ImageNet-64 RF++	2	1.083	2.495	1.097	2.521	25,386
ImageNet-64 RF++	4	1.081	2.487	1.094	2.561	24,990
ImageNet-64 RF++	8	1.071	2.490	1.092	2.522	25,376
ImageNet-64 EDM-VE→RF	2	1.114	2.607	1.132	2.619	24,433
ImageNet-64 EDM-VE→RF	4	1.123	2.583	1.156	2.630	24,338
ImageNet-64 EDM-VE→RF	8	1.127	2.586	1.146	2.648	24,169

Table 33: Text-to-image policy latency on NVIDIA H100, averaged over NFEs 4/6/8.

		Batch 1 latency (ms)		Batch 64 latency (ms)		Batch 64
Backbone	Policy	Policy	Complete	Policy	Complete	clocks/s
SANA	Deterministic	0.90	1.28	0.92	13.56	4,721
SANA	Stochastic	75.69	76.49	78.01	90.37	708
SD1.5	Deterministic	0.90	1.27	0.92	13.39	4,781
SD1.5	Stochastic	75.75	76.30	78.05	90.82	705

Table 34: Frozen-field latency on NVIDIA A100; 3D means use 43/43/42 members.

Dataset	NFE	Batch 1 (ms)	Batch 64 (ms)	Batch 64 samples/s
(a) Temporal sequences				
Solar	4	15.645	241.369	265.2
Solar	8	31.289	482.738	132.6
Solar	12	46.934	724.107	88.4
Solar	16	62.579	965.476	66.3
Traffic	4	15.821	31.644	2,022.5
Traffic	8	31.642	63.287	1,011.3
Traffic	12	47.462	94.931	674.2
Traffic	16	63.283	126.575	505.6
Weather	4	15.641	16.476	3,884.4
Weather	8	31.281	32.952	1,942.2
Weather	12	46.922	49.428	1,294.8
Weather	16	62.562	65.905	971.1
(b) Continuous 3D coordinates				
Set 1	4	15.632	17.363	3,686.1
Set 1	8	31.265	34.725	1,843.0
Set 1	12	46.897	52.088	1,228.7
Set 1	16	62.530	69.450	921.5
Set 2	4	15.685	17.403	3,677.6
Set 2	8	31.370	34.805	1,838.8
Set 2	12	47.055	52.208	1,225.9
Set 2	16	62.740	69.610	919.4
Set 3	4	15.624	17.199	3,721.1
Set 3	8	31.249	34.398	1,860.6
Set 3	12	46.873	51.597	1,240.4
Set 3	16	62.497	68.796	930.3

F.5 Matched qualitative examples

Fig. 7 shows matched examples; the corresponding complete numerical results are in Appx E.2, E.3, and E.4. The molecular examples include extended NFEs from Tab. 14.

G Text-to-Image Experiments

The COCO study compares deterministic GICO-global and GICO-prompt using the shared fitting and selection procedure (Appx C). ImageReward-v1.0 (IR) (Xu et al., 2023) and CLIP-FlanT5-XXL VQAScore (VQA) (Lin et al., 2025) supply fitting rewards and test metrics on disjoint prompts. VQA is a raw score, not benchmark accuracy.

G.1 Backbones, data, and evaluation protocol

Prompts come from COCO captions (Lin et al., 2014; Chen et al., 2015). A seeded caption is selected per source image before context sampling; image IDs define split groups, preventing related captions from crossing fitting, validation, and test boundaries.

Table 35: Text-to-image backbones and evaluation protocol.

Item	SANA experiment	SD1.5 experiment
Frozen backbone	SANA-600M-512px (Xie et al., 2025)	SD1.5 EMA (Rombach et al., 2022)
Output	512×512 pixels.	
Solver	Native FlowEuler	Official LD3 iPNDM, order 2
Guidance	CFG (Ho & Salimans, 2022) 4.5; PAG disabled	CFG 7.5
Main NFE	4, 6, 8.	
Main methods	Uniform; native; selected fixed; GICO-global; GICO-prompt; BO; clock PG	Uniform; native; selected fixed; GICO-global; GICO-prompt; LD3
Transfer	GICO fitted at 4/8, evaluated at 6.	
Controls	Utility weighting, nested 6/12/25 support, and surrogate top-1 at NFE 6	same
Prompt pools	Collection: 256 distinct source images, one caption each, split 204/52 for fitting/validation. Locked test: 512 disjoint prompts. Duplicate captions are excluded; observation allowances are in Tab. 36.	
Replication	Two matched generation seeds (17011, 94153); three optimizer seeds (3109, 6833, 104729).	
Completed test	252 runs; 104 distinct method/NFE/budget settings; 258,048 scored images.	

Generator, text encoder, VAE, and reward weights stay frozen. NFE counts generator field evaluations, not scorer calls or total time. SD1.5 native and uniform coincide under this iPNDM convention; SANA native uses a shifted schedule.

Reporting and uncertainty. Learned methods report the mean and sample SD of optimizer-run means; fixed clocks have no optimizer SD. Paired effects are raw score differences. Pointwise 95% intervals use 10,000 prompt-bootstrap resamples (seed 23119) after averaging generation and optimizer seeds. They condition on the fitted runs, do not resample optimizer seeds, and are unadjusted for multiplicity.

G.2 Reward definition and observation budgets

For $\ell \in \{\mathrm{IR}, \mathrm{VQA}\}$, let R_ℓ and $R_{\ell,\mathrm{uniform}}$ be paired scores averaged over generation repeats. Define

$$v_\ell = \frac{R_\ell - R_{\ell,\mathrm{uniform}}}{\sigma_\ell}, \qquad \widehat{S} = \frac{v_{\mathrm{IR}} + v_{\mathrm{VQA}}}{2 s_{\mathrm{cal}}}.$$

Each σ_ℓ is the population SD of nonuniform-minus-uniform score differences in eligible fitting evidence, balancing source NFEs. The scalar scale s_{cal} follows Appx D.2. Both are frozen before fitting and reused for validation and test; tables report raw IR/VQA.

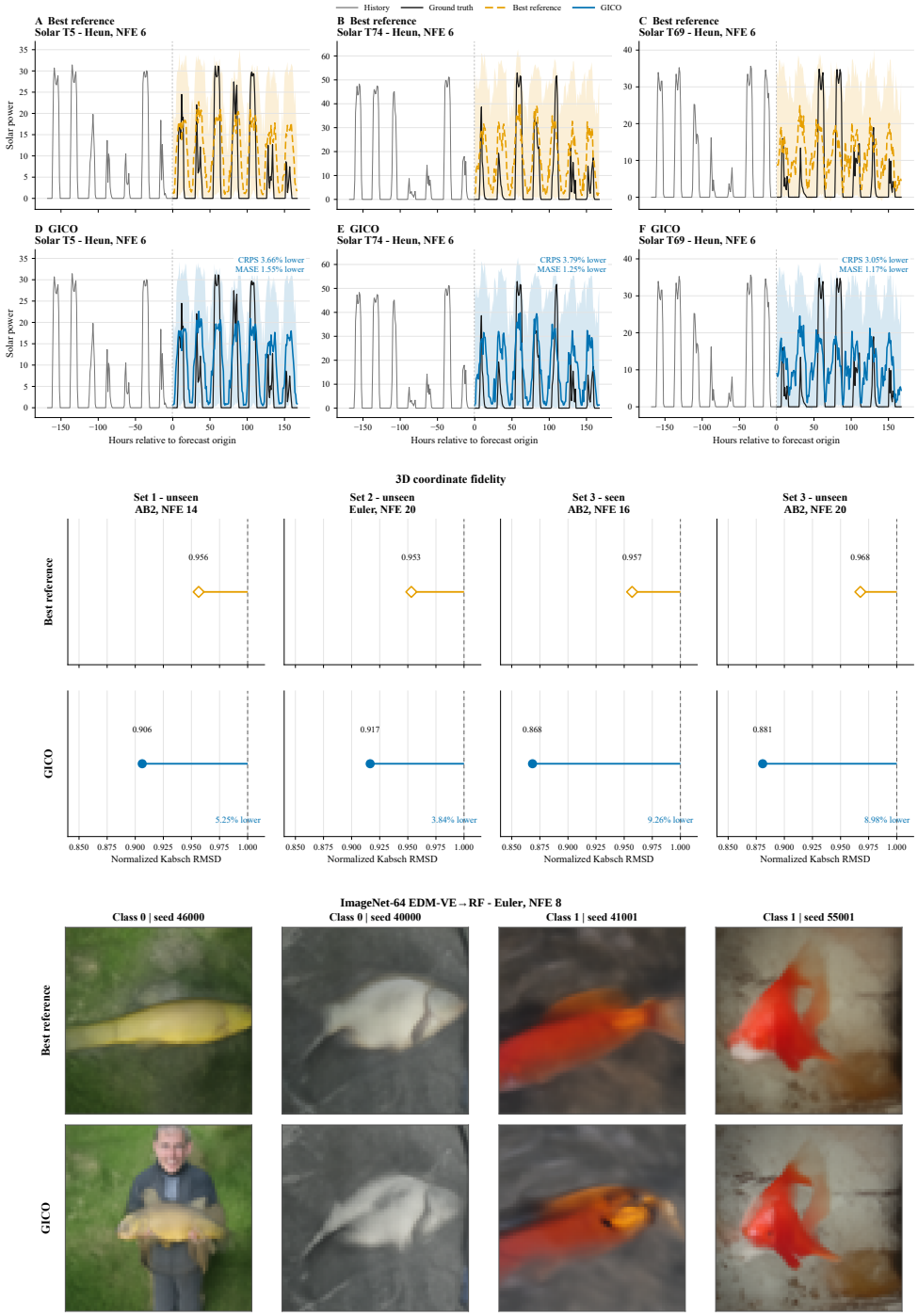

Figure 7: Matched comparisons: solar forecasts (Heun NFE 6), 3D coordinates (normalized Kabsch RMSD), and ImageNet-64 EDM-VE→RF (Euler NFE 8). Contexts and source seeds are matched; image columns share the class and noise seed.

Table 36: Observation allowances per optimizer run. Transfer combines the allowances at source NFEs 4 and 8.

Budget	Single-NFE allowance	4/8 transfer allowance
25%	1,600	3,200
50%	3,200	6,400
100%	6,400	12,800

GICO collects candidate–uniform pairs before optimization using the sparse assignments in Appx D.1. The collection manifest records the completed solves and context-disjoint fitting/validation split within each allowance. The 20% holdout reserves collected contexts and evidence, not new checkpoint solves. Uniform anchors count toward collection; surrogate predictions over the full reference pool are not additional observations. Each budget is fitted independently.

G.3 Baseline supervision and policy fitting

Tab. 37 contrasts offline fitting of fixed GICO evidence with adaptive utility collection by BO and clock PG. The comparison evaluates complete procedures, not update rules on identical observations. Clock PG uses KL coefficient 0.01 and entropy coefficient 0.001; Appx G.8 separately evaluates JS-REINFORCE.

Table 37: Baseline objectives and calibration access.

Method	Supervision / data access	Objective and learned object
Uniform / native	No calibration labels needed to define the prescribed schedule.	No optimization; prescribed clock.
Selected fixed	25 reference clocks evaluated on each calibration prompt with both noise seeds.	Select one reference maximizing mean $\widehat{S}$.
GICO-global	Collected candidate–uniform pairs; source-image-disjoint validation; zero prompt context.	Fit utility by ranking and Huber regression; optimize a 64-bin anchored policy, selected by held-out KL and surrogate utility.
GICO-prompt	Same collection and validation procedure; pooled frozen text embedding.	Same objectives and selection; prompt-conditioned deterministic density.
BO	25 full prompt-panel observations: uniform, Sobol initialization, adaptive proposals.	Maximize mean $\widehat{S}$ using SingleTaskGP, Matérn-5/2, qLogNEI (Balandat et al., 2020; Ament et al., 2023); select one global clock.
Clock PG	Uniform plus 24 prompt rounds; one sampled clock per prompt, scored with two noise seeds. Same representation as GICO-prompt.	Maximize terminal $\widehat{S}$ with a PPO-style loss (Schulman et al., 2017), KL penalty and entropy bonus; learn a logistic-normal clock policy.
LD3	100 teacher trajectories / fit: 50 calibration prompts $\times$ two seeds; 25 training and 25 holdout prompts. Teacher uses official GITS with one more field evaluation than the student.	Minimize LPIPS (Zhang et al., 2018) to generated teacher images; learn discretization parameters with separate integration/evaluation times and official additional freedoms.

Observation allowances do not equate fitting compute or supervision. Clock PG applies DDPO-inspired updates (Black et al., 2024) to whole clocks. LD3 matches inference NFE only; its generated LPIPS targets, fitting budget, and discretization freedoms differ.

G.4 Absolute results and calibration-selected references

Tabs. 39 and 40 report the independent observation budgets. SANA BO selects uniform at NFE 8 in every run after all 25 panels, explaining its zero SD. Fixed references maximize calibration $\widehat{S}$, never test metrics (Tab. 38). Fig. 8 shows optimizer-run means and prompt uncertainty.

Table 38: Calibration-selected fixed clocks. Late-p abbreviates its exponent; R denotes grid reversal. AYS and OTS are transferred references: "native" denotes their source coordinate, not the target backbone's prescribed clock.

Backbone	NFE	25%	50%	100%
SANA	4	Late-2 (R)	Late-2 (R)	Late-1.5 (R)
SANA	6	Late-2 (R)	Late-2 (R)	Late-1.5 (R)
SANA	8	Late-3 (R)	Late-1.5 (R)	Late-3 (R)
SD1.5	4	AYS native (R)	Uniform	Uniform
SD1.5	6	OTS native (R)	Late-1.5 (R)	Late-1.5 (R)
SD1.5	8	Late-1.5 (R)	Late-1.5 (R)	OTS native

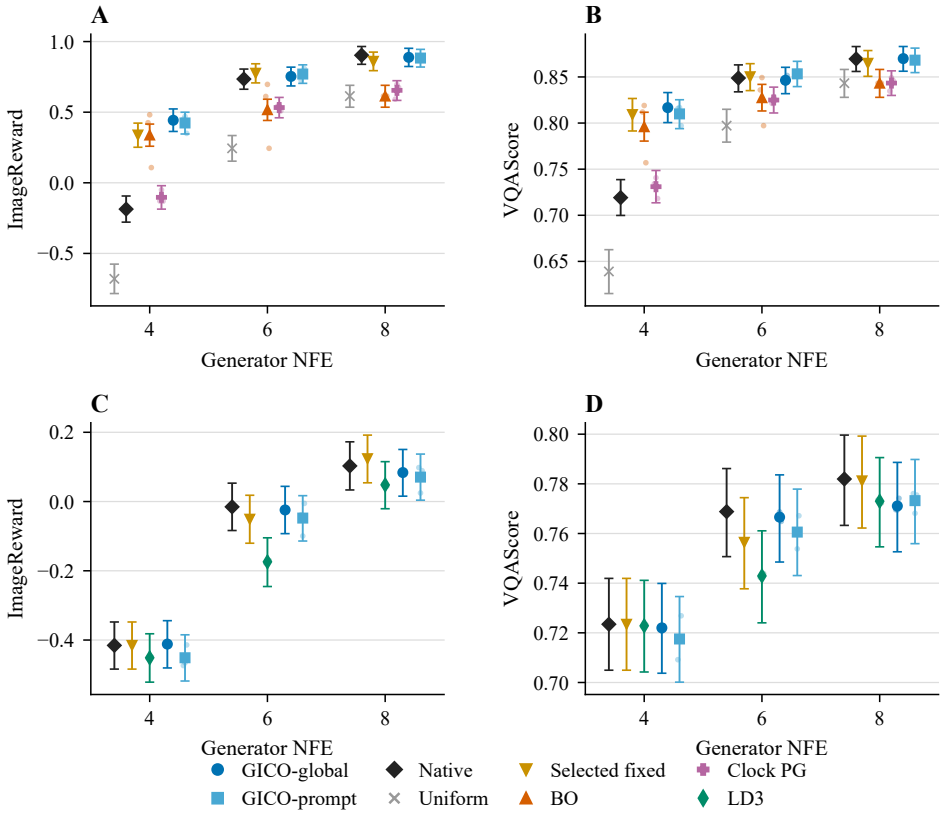

Figure 8: Full-budget text-to-image scores. **A/B**: SANA IR/VQA; **C/D**: SD1.5 IR/VQA. Large points average optimizer runs; small dots show individual run means. Bars follow Appx G.1.

Table 39: SANA absolute results across observation budgets. Budget cells list IR above VQA (both higher-is-better); fitted methods include optimizer SDs. Fixed IR / VQA pairs are shown once per NFE. All fitted rows target $\widehat{S}$ with method-specific objectives.

NFE	Method	25%	50%	100%
4	Uniform	−0.6799 / 0.6391 (all budgets)		
4	Native	−0.1864 / 0.7191 (all budgets)		
4	Selected fixed	0.3759 0.8062	0.3759 0.8062	0.3384 0.8092
4	GICO-global	0.4743 ± 0.0006 0.8157 ± 0.0016	0.4786 ± 0.0019 0.8158 ± 0.0009	0.4429 ± 0.0069 0.8167 ± 0.0011

Table 39: Continued.

NFE	Method	25%	50%	100%
		0.2843 ± 0.0229	0.3286 ± 0.0352	0.4236 ± 0.0644
4	GICO-prompt	0.7848 ± 0.0049	0.7958 ± 0.0018	0.8099 ± 0.0111
		0.4475 ± 0.0818	0.3579 ± 0.1615	0.3382 ± 0.2016
4	BO	0.8158 ± 0.0093	0.8070 ± 0.0188	0.7960 ± 0.0340
		−0.2786 ± 0.1111	−0.1684 ± 0.0552	−0.1028 ± 0.0458
4	Clock PG	0.6973 ± 0.0250	0.7187 ± 0.0064	0.7310 ± 0.0116
6	Uniform	0.2441 / 0.7972 (all budgets)		
6	Native	0.7347 / 0.8488 (all budgets)		
		0.8120	0.8120	0.7760
6	Selected fixed	0.8583	0.8583	0.8502
		0.7847 ± 0.0178	0.7610 ± 0.0174	0.7531 ± 0.0130
6	GICO-global	0.8547 ± 0.0026	0.8479 ± 0.0036	0.8464 ± 0.0045
		0.6883 ± 0.0224	0.7202 ± 0.1344	0.7700 ± 0.0421
6	GICO-prompt	0.8439 ± 0.0051	0.8466 ± 0.0117	0.8534 ± 0.0029
		0.6078 ± 0.1203	0.5239 ± 0.2424	0.5174 ± 0.2405
6	BO	0.8394 ± 0.0091	0.8216 ± 0.0211	0.8275 ± 0.0271
		0.3568 ± 0.0189	0.4293 ± 0.0315	0.5328 ± 0.0205
6	Clock PG	0.8021 ± 0.0037	0.8107 ± 0.0062	0.8250 ± 0.0036
8	Uniform	0.6140 / 0.8432 (all budgets)		
8	Native	0.8830 / 0.8696 (all budgets)		
		0.8603	0.8625	0.8603
8	Selected fixed	0.8648	0.8613	0.8648
		0.8910 ± 0.0037	0.8905 ± 0.0005	0.8883 ± 0.0077
8	GICO-global	0.8718 ± 0.0010	0.8678 ± 0.0006	0.8699 ± 0.0020
		0.8231 ± 0.0229	0.8397 ± 0.0454	0.8831 ± 0.0058
8	GICO-prompt	0.8603 ± 0.0029	0.8618 ± 0.0050	0.8682 ± 0.0013
		0.6140 ± 0.0000	0.6140 ± 0.0000	0.6140 ± 0.0000
8	BO	0.8432 ± 0.0000	0.8432 ± 0.0000	0.8432 ± 0.0000
		0.5964 ± 0.0206	0.6787 ± 0.0584	0.6546 ± 0.0550
8	Clock PG	0.8345 ± 0.0030	0.8447 ± 0.0085	0.8434 ± 0.0077

Table 40: SD1.5 absolute results across observation budgets. Budget cells list IR above VQA (both higher-is-better); fitted methods include optimizer SDs. Fixed IR / VQA pairs are shown once per NFE. LD3 uses its separate fitting budget, not a 25/50/100% allowance.

NFE	Method	25%	50%	100%
4	Native / uniform	−0.4155 / 0.7235 (all budgets)		
		−0.4998	−0.4155	−0.4155
4	Selected fixed	0.7023	0.7235	0.7235
		−0.4202 ± 0.0053	−0.4182 ± 0.0056	−0.4117 ± 0.0017
4	GICO-global	0.7218 ± 0.0008	0.7216 ± 0.0003	0.7220 ± 0.0011
		−0.4517 ± 0.0227	−0.4727 ± 0.0209	−0.4212 ± 0.0323
4	GICO-prompt	0.7193 ± 0.0060	0.7176 ± 0.0048	0.7175 ± 0.0089
4	LD3	−0.4513 ± 0.0022 / 0.7228 ± 0.0004 (LPIPS fit)		
6	Native / uniform	−0.0153 / 0.7688 (all budgets)		
		−0.1026	−0.0510	−0.0510
6	Selected fixed	0.7459	0.7565	0.7565
		−0.0298 ± 0.0101	−0.0367 ± 0.0113	−0.0243 ± 0.0024
6	GICO-global	0.7684 ± 0.0009	0.7682 ± 0.0003	0.7666 ± 0.0019
		−0.1246 ± 0.1028	−0.0472 ± 0.0370	−0.0478 ± 0.0478
6	GICO-prompt	0.7572 ± 0.0105	0.7627 ± 0.0082	0.7606 ± 0.0067
6	LD3	−0.1741 ± 0.0038 / 0.7429 ± 0.0008 (LPIPS fit)		
8	Native / uniform	0.1028 / 0.7819 (all budgets)		
		0.0237	0.0237	0.1238
8	Selected fixed	0.7570	0.7570	0.7812
		0.1011 ± 0.0077	0.0809 ± 0.0084	0.1837 ± 0.0011
8	GICO-global	0.7770 ± 0.0019	0.7701 ± 0.0037	0.7711 ± 0.0027
		0.0818 ± 0.0059	0.0454 ± 0.0417	0.1705 ± 0.0400
8	GICO-prompt	0.7782 ± 0.0004	0.7729 ± 0.0051	0.7732 ± 0.0044
8	LD3	0.0481 ± 0.0032 / 0.7730 ± 0.0006 (LPIPS fit)		

G.5 PAIRED COMPARISONS

Tabs. 41–42 give the paired contrasts discussed in Sec. 6.1, using the interval convention in Appx G.1. Fig. 9 extends the BO/PG comparison across observation budgets; baseline matching is specified in Tab. 37.

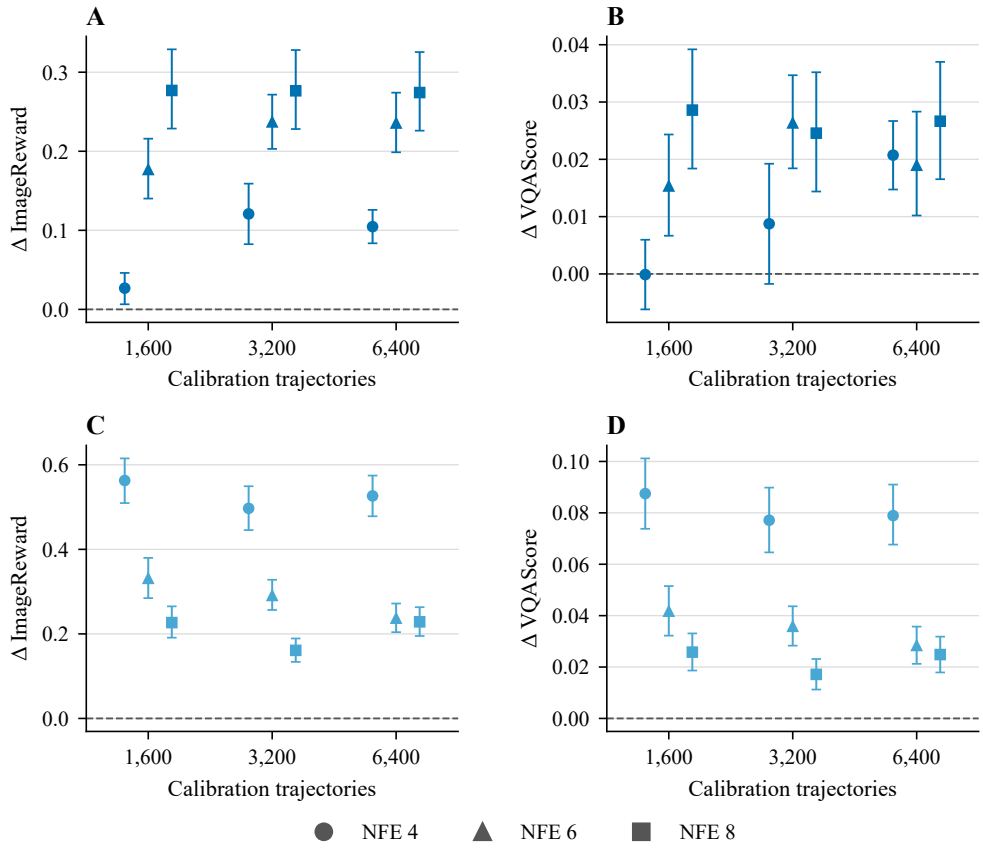

Figure 9: SANA observation-budget comparisons. **A/B**: global GICO minus BO, IR/VQA; **C/D**: prompt GICO minus clock PG, IR/VQA. Positive favors GICO. Independent observation allowances include GICO selection; supervision and compute differ.

Table 41: SANA paired differences at full budget: mean [95% prompt interval].

NFE	Candidate minus baseline	Δ IR	Δ VQA
4	GICO-global − Native	+0.6292 [+0.5747, +0.6842]	+0.0976 [+0.0823, +0.1130]
4	GICO-global − Selected fixed	+0.1044 [+0.0711, +0.1386]	+0.0075 [−0.0019, +0.0174]
4	GICO-prompt − Selected fixed	+0.0851 [+0.0424, +0.1298]	+0.0007 [−0.0104, +0.0124]
4	BO − GICO-global	−0.1046 [−0.1258, −0.0836]	−0.0207 [−0.0267, −0.0147]
4	Clock PG − GICO-prompt	−0.5264 [−0.5746, −0.4782]	−0.0789 [−0.0910, −0.0676]
6	GICO-global − Native	+0.0184 [−0.0140, +0.0510]	−0.0024 [−0.0111, +0.0063]
6	GICO-global − Selected fixed	−0.0230 [−0.0515, +0.0056]	−0.0038 [−0.0114, +0.0038]
6	GICO-prompt − Selected fixed	−0.0060 [−0.0334, +0.0221]	+0.0031 [−0.0038, +0.0103]
6	BO − GICO-global	−0.2357 [−0.2741, −0.1987]	−0.0190 [−0.0283, −0.0102]
6	Clock PG − GICO-prompt	−0.2372 [−0.2719, −0.2041]	−0.0284 [−0.0357, −0.0212]
8	GICO-global − Native	+0.0053 [−0.0156, +0.0256]	+0.0003 [−0.0059, +0.0064]
8	GICO-global − Selected fixed	+0.0280 [+0.0154, +0.0410]	+0.0051 [+0.0010, +0.0092]
8	GICO-prompt − Selected fixed	+0.0228 [+0.0054, +0.0405]	+0.0034 [−0.0020, +0.0086]

Table 41: Continued.

NFE	Candidate minus baseline	Δ IR	Δ VQA
8	BO − GICO-global	-0.2743 $[-0.3255, -0.2260]$	-0.0267 $[-0.0370, -0.0165]$
8	Clock PG − GICO-prompt	-0.2285 $[-0.2632, -0.1950]$	-0.0249 $[-0.0318, -0.0179]$

Table 42: SD1.5 paired differences at full budget: mean [95% prompt interval].

NFE	Candidate minus baseline	Δ IR	Δ VQA
4	GICO-global − Native	$+0.0038$ $[-0.0082, +0.0156]$	-0.0015 $[-0.0064, +0.0036]$
4	GICO-prompt − Native	-0.0057 $[-0.0385, +0.0267]$	-0.0059 $[-0.0159, +0.0040]$
4	LD3 − Native	-0.0358 $[-0.0632, -0.0074]$	-0.0006 $[-0.0104, +0.0094]$
4	GICO-global − LD3*	$+0.0396$ $[+0.0121, +0.0671]$	-0.0008 $[-0.0104, +0.0087]$
6	GICO-global − Native	-0.0090 $[-0.0337, +0.0154]$	-0.0022 $[-0.0096, +0.0054]$
6	GICO-prompt − Native	-0.0325 $[-0.0579, -0.0062]$	-0.0082 $[-0.0159, -0.0005]$
6	LD3 − Native	-0.1588 $[-0.1933, -0.1251]$	-0.0259 $[-0.0360, -0.0156]$
6	GICO-global − LD3*	$+0.1498$ $[+0.1162, +0.1839]$	$+0.0237$ $[+0.0142, +0.0327]$
8	GICO-global − Native	$+0.0809$ $[+0.0482, +0.1132]$	-0.0109 $[-0.0195, -0.0023]$
8	GICO-prompt − Native	$+0.0677$ $[+0.0435, +0.0918]$	-0.0087 $[-0.0154, -0.0021]$
8	LD3 − Native	-0.0547 $[-0.0888, -0.0211]$	-0.0089 $[-0.0189, +0.0012]$
8	GICO-global − LD3*	$+0.1356$ $[+0.0997, +0.1713]$	-0.0019 $[-0.0117, +0.0078]$

*Additional post hoc contrasts of already-frozen outputs. LD3 uses LPIPS and intervals follow Appx G.1.

G.6 HELD-OUT-NFE TRANSFER

Transfer deploys zero-shot at NFE 6 after fitting at 4/8; NFE 6 is excluded from fitting, normalization, and selection. Allowances apply per source NFE (Tab. 36): full transfer has 12,800 observations, versus 6,400 for direct fitting. Tabs. 43–44 report scores and post hoc contrasts; Fig. 10 includes reduced budgets.

Table 43: Held-out NFE 6 after fitting at NFEs 4/8; mean ± optimizer SD.

Backbone	Budget	Method	Fit reward	IR ↑	VQA ↑
SANA	25%	GICO-global	$\widehat{S}$	0.7998 ± 0.0124	0.8548 ± 0.0035
SANA	25%	GICO-prompt	$\widehat{S}$	0.6780 ± 0.1196	0.8434 ± 0.0130
SANA	50%	GICO-global	$\widehat{S}$	0.7544 ± 0.0117	0.8454 ± 0.0041
SANA	50%	GICO-prompt	$\widehat{S}$	0.7428 ± 0.0620	0.8448 ± 0.0083
SANA	100%	GICO-global	$\widehat{S}$	0.7746 ± 0.0136	0.8502 ± 0.0038
SANA	100%	GICO-prompt	$\widehat{S}$	0.7949 ± 0.0169	0.8526 ± 0.0046
SD1.5	25%	GICO-global	$\widehat{S}$	-0.0202 ± 0.0121	0.7684 ± 0.0028
SD1.5	25%	GICO-prompt	$\widehat{S}$	-0.0811 ± 0.0304	0.7608 ± 0.0023
SD1.5	50%	GICO-global	$\widehat{S}$	-0.0326 ± 0.0099	0.7638 ± 0.0043
SD1.5	50%	GICO-prompt	$\widehat{S}$	-0.0815 ± 0.0587	0.7602 ± 0.0034
SD1.5	100%	GICO-global	$\widehat{S}$	-0.0230 ± 0.0075	0.7663 ± 0.0016
SD1.5	100%	GICO-prompt	$\widehat{S}$	-0.0358 ± 0.0204	0.7608 ± 0.0027

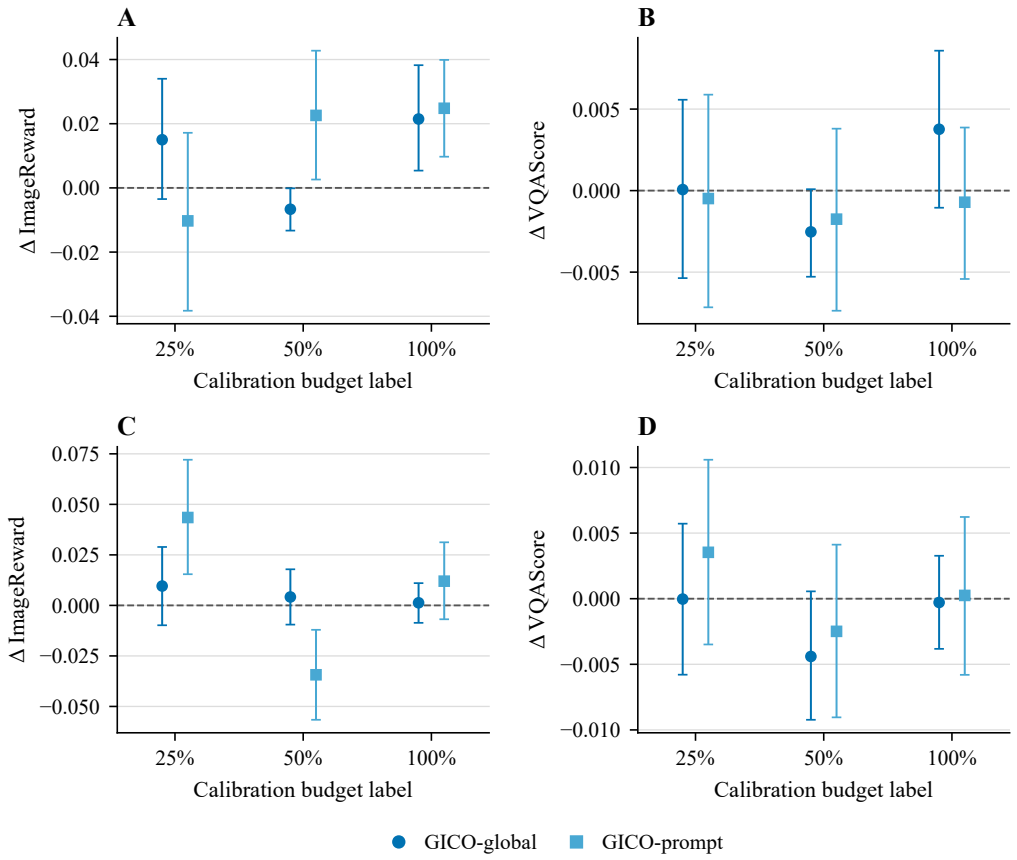

Figure 10: Transfer minus direct fitting at NFE 6. **A/B**: SANA IR/VQA; **C/D**: SD1.5 IR/VQA. Transfer fits at 4/8 with twice the total observations.

Table 44: Post hoc transfer-minus-direct differences at NFE 6, full budget.

NFE	Candidate minus baseline	Δ IR	Δ VQA
6	SANA global: transfer − direct	$+0.0215$ $[+0.0054, +0.0382]$	$+0.0038$ $[-0.0011, +0.0086]$
6	SANA prompt: transfer − direct	$+0.0248$ $[+0.0097, +0.0399]$	-0.0007 $[-0.0054, +0.0039]$
6	SD1.5 global: transfer − direct	$+0.0013$ $[-0.0086, +0.0110]$	-0.0003 $[-0.0038, +0.0033]$
6	SD1.5 prompt: transfer − direct	$+0.0120$ $[-0.0069, +0.0312]$	$+0.0003$ $[-0.0058, +0.0062]$

G.7 Utility, support, and surrogate controls

Controls use SANA at NFE 6 and the paired GICO-prompt baseline in Tab. 45. IR/VQA controls change scalar utility weights while retaining both regressed components. Nested 6/12/25-clock supports have observation allowances of 1,536/3,072/6,400, with the same collection and selection rules. Surrogate top-1 selects the highest-scored reference instead of decoding the density policy.

IR-only weighting raises IR by 0.0236, whereas VQA-only weighting lowers both scores. Six-clock support and surrogate top-1 reduce IR by 0.0233 and 0.0481. Their VQA intervals, and both intervals for 12-clock support, include zero (Tab. 46); these intervals do not establish equivalence. The controls distinguish utility weighting, reference coverage, and density learning from discrete reference selection.

Table 45: SANA utility, support, and surrogate controls at NFE 6.

Method / control	Utility	Support	IR ↑	VQA ↑
GICO-prompt baseline	Balanced $\widehat{S}$	25	0.7647 ± 0.0282	0.8540 ± 0.0023
IR-utility control	$\widehat{u}_{\mathrm{IR}}$	25	0.7884 ± 0.0151	0.8530 ± 0.0025
VQA-utility control	$\widehat{u}_{\mathrm{VQA}}$	25	0.7052 ± 0.0937	0.8449 ± 0.0102
6-clock support	Balanced $\widehat{S}$	6	0.7415 ± 0.0086	0.8526 ± 0.0015
12-clock support	Balanced $\widehat{S}$	12	0.7647 ± 0.0229	0.8573 ± 0.0036
Surrogate top-1	Balanced $\widehat{S}$	25	0.7166 ± 0.0446	0.8518 ± 0.0036

Table 46: Control-minus-GICO-prompt differences: mean [95% prompt interval].

NFE	Candidate minus baseline	Δ IR	Δ VQA
6	IR-utility control − GICO-prompt	$+0.0236$ $[+0.0054, +0.0422]$	-0.0010 $[-0.0066, +0.0046]$
6	VQA-utility control − GICO-prompt	-0.0595 $[-0.0840, -0.0358]$	-0.0091 $[-0.0154, -0.0031]$
6	6-clock support − GICO-prompt	-0.0233 $[-0.0458, -0.0011]$	-0.0014 $[-0.0070, +0.0042]$
6	12-clock support − GICO-prompt	$+0.0000$ $[-0.0220, +0.0218]$	$+0.0033$ $[-0.0021, +0.0089]$
6	Surrogate top-1 − GICO-prompt	-0.0481 $[-0.0730, -0.0231]$	-0.0022 $[-0.0082, +0.0039]$

G.8 Additional James–Stein clock-policy comparison

Tab. 47 compares GICO-global with our paper-based JS-REINFORCE reconstruction (Yu et al., 2026) on SANA-600M at 512×512, with a 3,200-observation allowance per method. Comparisons use this table's matched uniform baseline. JS samples a joint Dirichlet schedule, uses two rollouts per context, and applies REINFORCE with a James–Stein-shrunken leave-one-out baseline, without PPO clipping or KL/entropy regularization. Its head conditions on initial Gaussian image noise and frozen text features; GICO-global uses neither. This compares complete methods rather than isolating their optimizers.

For b field evaluations, the JS action defines $b + 1$ intervals, supplies b interior evaluation times, and appends the terminal endpoint for b Euler updates. Uniform anchors and sampled trajectories count toward its allowance. GICO uses collected pairs and the generator-free selector in Appx C.2. GICO has higher mean IR at all three NFEs and higher mean VQA at 6/8; JS has higher mean VQA at NFE 4. The table reports point estimates.

Table 47: Additional JS-REINFORCE policy comparison on SANA-600M at 512×512 resolution with a 3,200-observation allowance per method. ImageReward (IR) and VQAScore (VQA) are higher-is-better.

	NFE 4		NFE 6		NFE 8	
Method	IR ↑	VQA ↑	IR ↑	VQA ↑	IR ↑	VQA ↑
Uniform	-0.7697	0.6676	0.1025	0.8193	0.4822	0.8669
JS-REINFORCE	0.1510	$\mathbf{0.8274}$	0.3733	0.8355	0.5328	0.8652
GICO-global	$\mathbf{0.5789}$	0.8211	$\mathbf{0.6128}$	$\mathbf{0.8678}$	$\mathbf{0.7169}$	$\mathbf{0.8777}$

G.9 Computation and accounting

Tab. 48 records study-wide generation and scorer costs across both backbones, separately from per-fit observation allowances. These are measurement-phase totals, not surrogate/policy optimization costs. GICO fitting and checkpoint selection use saved evidence and surrogate predictions; generator-backed validation reports do not alter the selected policy. The separate JS comparison is accounted independently.

Table 48: Instrumented generation and reward-scoring costs across both latent models.

Phase	Completed trajectories	Generation GPU-h	IR + VQA GPU-h
calibration	231,936	23.674	8.942
locked_test	258,048	40.369	10.173
pilot	768	0.138	0.031
validation	64,512	10.118	2.568

G.10 Matched qualitative examples

Score-independent prompt selection uses a fixed-prefix ordering of prompt IDs. Fig. 11 extends Fig. 2 with all methods for "A dog runs along a path carrying a frisbee in its mouth." and "this is a market that sells fresh fruit". All images are uncropped 512-pixel outputs with generation seed 17011 and optimizer seed 3109. SD1.5 native/uniform outputs coincide and share a row.

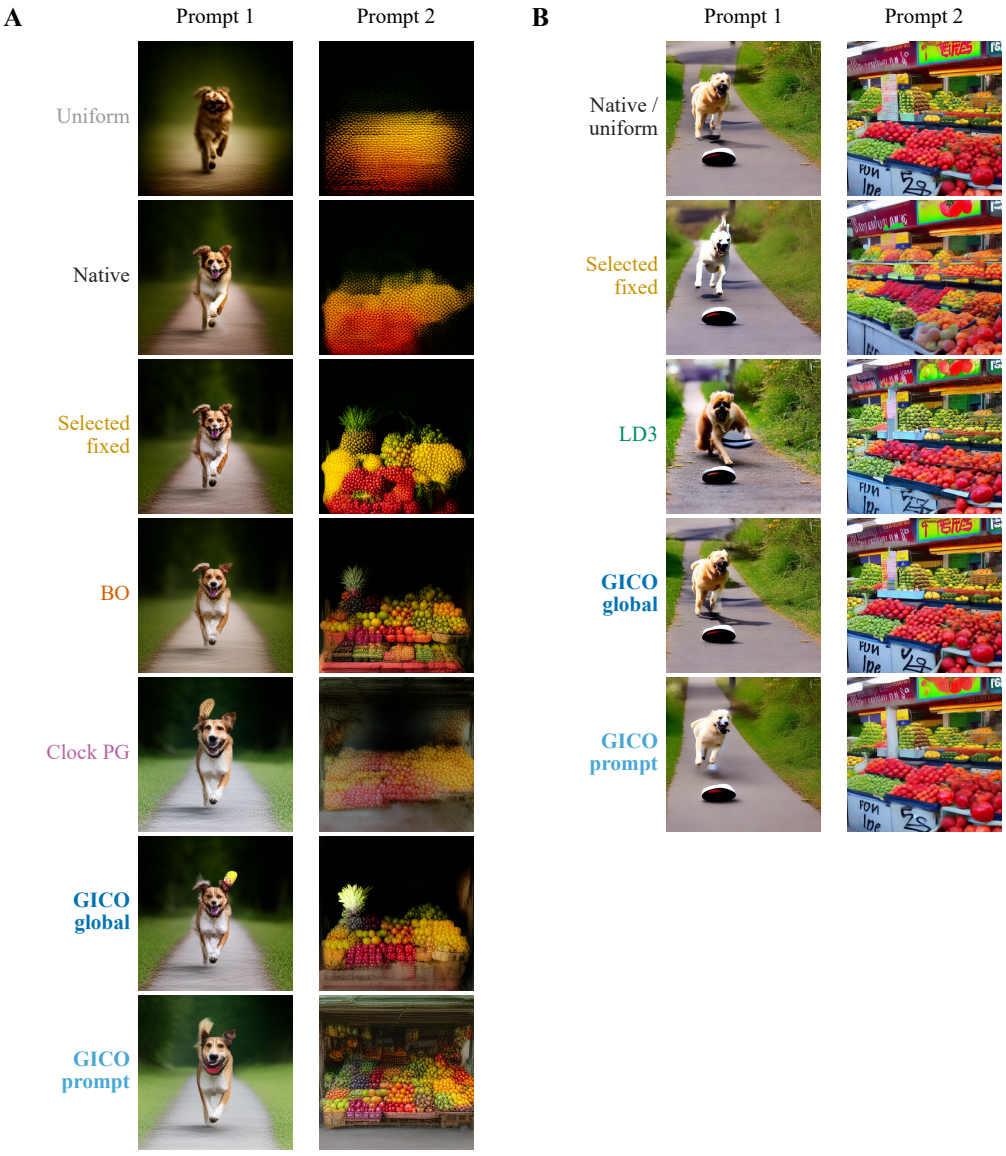

Figure 11: Full gallery of Fig. 2. Rows use the same two prompts and noise.

G.11 REFERENCE-HULL AND DECODED-GRID DIAGNOSTICS

For each backbone/solver/NFE setting, let $\mathcal{H}_\xi = \mathrm{conv}\{\boldsymbol{p}_i\}_{i=1}^{25}$ be the hull of the guarded 64-bin reference masses (Appx B.1). The hull distance is

$$d_{\mathrm{TV}}(\boldsymbol{p}, \mathcal{H}_\xi) = \min_{\boldsymbol{w}\in\Delta_{25}} \frac{1}{2} \left\| \boldsymbol{p} - \sum_{i=1}^{25} w_i \boldsymbol{p}_i \right\|_1 .$$

For policy nodes $\tau_k = F^{-1}_{\rho_{\boldsymbol{p}}}(k/K)$, an exact in-hull match exists iff some $\boldsymbol{w} \in \Delta_{25}$ satisfies $\sum_i w_i F_{\rho_{\boldsymbol{p}_i}}(\tau_k) = k/K$, $k = 1, \ldots, K-1$. This linear feasibility test searches the full hull, not just the nearest-TV projection. We classify densities as outside the hull when TV exceeds 10^{-7}, use LP feasibility tolerance 10^{-9}, and verify returned mixtures by maximum decoded-node error below 10^{-7} on normalized time $[0, 1]$.

Table 49: Reference-hull diagnostics. Top: six audited deterministic clocks. Bottom: 768 stochastic clocks per row, all outside the hull; "Prompt" means prompt-conditioned. Brackets give p10–p90 over sampled clocks, not confidence intervals. Node match reports the stated feasibility/verification test.

Deterministic policies

Backbone / solver	NFE	Hull TV	In-hull node match
SANA / FlowEuler	4	0.008432	No
	6	0.006952	No
	8	0.006328	Yes
SD1.5 / iPNDM-2	4	0.002117	No
	6	0.002971	Yes
	8	0.035513	No

Stochastic policies

Backbone / solver	NFE	Policy	Median hull TV [p10–p90]	Node match
SANA / FlowEuler	4	Global	0.0730 [0.0580, 0.0934]	52.9%
	4	Prompt	0.0763 [0.0602, 0.0972]	37.2%
	6	Global	0.0612 [0.0482, 0.0788]	64.1%
	6	Prompt	0.0645 [0.0506, 0.0821]	47.8%
	8	Global	0.0548 [0.0425, 0.0712]	81.3%
	8	Prompt	0.0581 [0.0449, 0.0746]	65.5%
SD1.5 / iPNDM-2	4	Global	0.0245 [0.0178, 0.0345]	66.8%
	4	Prompt	0.0278 [0.0199, 0.0382]	53.4%
	6	Global	0.0312 [0.0226, 0.0431]	43.1%
	6	Prompt	0.0346 [0.0248, 0.0470]	59.5%
	8	Global	0.0584 [0.0195, 0.1140]	22.1%
	8	Prompt	0.0712 [0.0280, 0.1315]	13.8%

All six deterministic densities lie outside the reference hull, only two grids have an in-hull match. Density-hull departure and decoded-grid reproducibility are therefore distinct.