PATCH THE PATH: FIXED-BUDGET POST-SAMPLING TRAJECTORY UPGRADE FOR FLOW MATCHING

Anonymous authors
Paper under double-blind review

ABSTRACT

We study how to improve a flow-matching sample after sampling is complete, using a small, fixed budget of additional model evaluations. We propose FixUp-FM, a training-free method that revisits the completed Euler trajectory and upgrades steps with large estimated local errors to Heun. A minimum-norm velocity adjustment derived from the trajectory propagates these corrections downstream without additional model evaluations, reserving computation for selected Heun corrections. FixUp-FM also supports incremental refinement across successive requests, with each pass building on the latest corrected trajectory rather than restarting. Under smoothness and bounded-propagation assumptions in the fixed-budget setting, we establish near-oracle defect selection and second-order endpoint agreement with a selective Euler–Heun rerun using exact velocities. Across SD3 Medium, SD3.5 Large, and FLUX.1-dev, FixUp-FM achieves higher mean ImageReward than NFE-matched direct Euler in 111 of 120 tested correction-budget configurations across the main sampling range.

1 INTRODUCTION

Flow-matching models generate images by numerically integrating a learned velocity field, with sampling cost measured in neural function evaluations (NFEs). In many practical settings, an image is first generated using a standard sampler, and the user then decides whether to allocate additional computation for further improvement. At that point, since the initial computation is already a sunk cost, the question is how to build on the completed run to improve the sample rather than start over.

We formulate this setting as *fixed-budget post-sampling trajectory correction*. Given a completed a-NFE Euler run and an additional budget of b NFEs, we ask how to improve the resulting sample by reusing information from the completed trajectory, ideally matching or exceeding the quality of a fresh $(a + b)$-NFE run. Here, we target the completed numerical trajectory, rather than modify the initial sampling procedure or apply a separate image-refinement stage.

This completed trajectory provides more than a final image. Its velocities allow local Euler errors to be estimated, identifying the steps with the largest errors and providing information useful for their correction. The key challenge is that these same velocities are tied to the original trajectory. Once a step is corrected, downstream states shift, and the original velocities no longer correspond to them. Yet re-evaluating the model at every affected state could require recomputing much of the remaining trajectory, approaching the cost of a fresh run.

We propose **FixUp-FM**, a training-free method that upgrades selected Euler steps to Heun and propagates their corrections downstream. Our key observation is that neighboring states and velocities reveal a local directional relation between state and velocity changes. We extend this relation with a minimum-norm linear map, allowing downstream velocities to be adjusted as corrections move the trajectory without re-evaluating the model. New model evaluations are therefore concentrated on the selected Heun stages. Together with exact reuse of the first selected endpoint stage, this reduces the additional cost of upgrading K steps to $K - 1$ NFEs, except in the terminal-only edge case. FixUp-FM can also be repeated as further budget becomes available, with successive passes building on the latest corrected trajectory and progressively improving the same sample.

Experiments on SD3 Medium, SD3.5 Large, and FLUX.1-dev (Esser et al., 2024; Stability AI, 2024; Black Forest Labs, 2024) show higher mean ImageReward (Xu et al., 2023a) than NFE-matched

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.

direct Euler at 20 of the 21 main operating points. Notably, starting from Euler-12 outputs, a single correction improves ImageReward by 0.050 and 0.058 on SD3 and SD3.5, respectively, with zero additional NFEs. Under successive correction requests, FixUp-FM remains above both the initial Euler output and NFE-matched direct Euler throughout the evaluated continuation schedules.

Our contributions are threefold. **(1) Fixed-budget post-sampling correction.** We formulate the problem of improving a completed Euler trajectory under a limited additional NFE budget and introduce FixUp-FM, a training-free framework that allocates the added budget to selective trajectory upgrades, either in a single pass or over successive passes. **(2) Minimum-norm directional propagation.** We derive a minimum-norm velocity adjustment from directional information identified by the reference trace, avoiding model re-evaluation at every downstream state. Under locally affine dynamics, it exactly recovers the velocity for the initial correction at any interior velocity node. **(3) End-to-end numerical guarantees.** We establish near-oracle defect selection and end-to-end correction guarantees. Under smoothness and bounded-propagation assumptions in the fixed-budget setting, FixUp-FM achieves second-order agreement with an ideal selective Euler–Heun rerun using exact velocities, while successive requests remain consistent with the one-shot correction.

2 Related work

Numerical sampling methods. Existing approaches improve generation through higher-order integration, correction, or iterative refinement. UniPC uses a corrector to increase the order of accuracy without extra model evaluations (Zhao et al., 2023), while DC-Solver dynamically compensates for corrector-induced state–output misalignment (Zhao et al., 2024). ERK-Guid uses solver-error estimates to guide diffusion sampling in stiff regions (Kong et al., 2026). SRDS uses Parareal to refine a coarse few-step solve toward the serial solution, trading extra parallel compute for lower sampling latency (Selvam et al., 2024). By contrast, FixUp-FM leaves the initial Euler pass unchanged and corrects the completed trajectory under a small, fixed additional NFE budget.

Image refinement and resampling. Some methods refine samples through dedicated refinement stages or restructured denoising processes. SDXL applies a separate denoising refiner (Podell et al., 2024), while Peng & Gao (2026) learn a separate data-space refinement flow. SDEdit re-noises input images before denoising them (Meng et al., 2022), while Restart repeatedly alternates re-noising with backward ODE integration (Xu et al., 2023b). Nested Diffusion embeds an inner sampler at each outer step, enabling anytime generation (Elata et al., 2024). FixUp-FM instead revisits the completed Euler trajectory using the original velocity field, without a separate refiner or re-noising.

Noise optimization and candidate selection. D-Flow optimizes the source noise by differentiating through the generative flow for controlled generation (Ben-Hamu et al., 2024), while ReNO optimizes the noise of one-step models using human-preference rewards (Eyring et al., 2024). Ma et al. (2025) use verifier scores to select among candidate noise samples and sampling paths. Rollover Budget Forcing allocates computation across timesteps for reward-guided particle sampling in flow models (Kim et al., 2025). ABCD repeatedly re-noises and denoises selected candidates, with reward-based selection and adaptive stopping (Lee et al., 2025). Rather than optimizing the initial noise or searching over alternatives, FixUp-FM allocates its additional evaluations to correcting the completed trajectory.

3 Method: FixUp-FM

Given a completed Euler trajectory and a budget of additional NFEs, we seek to reduce its numerical error without restarting sampling. FixUp-FM selects intervals with large estimated local errors, upgrades them to Heun, and propagates the resulting state changes downstream. The key is to use the completed Euler trajectory both to **locate these errors** and to **estimate how a state displacement changes the velocity**, reserving new model evaluations for selected Heun stages.

Formally, for a fixed condition c, flow-matching sampling follows $\dot{x}(t) = f_\theta(x(t), t; c)$ from prior noise $x(0) \sim p_0$ on $[0, 1]$. Write $f(x, t) = f_\theta(x, t; c)$ and use the grid $0 = t_0 < \cdots < t_a = 1$, with $a \geq 2$. Interval k is $[t_k, t_{k+1}]$ with step size $h_k = t_{k+1} - t_k$. An a-NFE Euler sampler produces

$$\bar{x}_{k+1} = \bar{x}_k + h_k \bar{v}_k, \qquad \bar{v}_k = f(\bar{x}_k, t_k), \qquad k = 0, \ldots, a-1. \tag{1}$$

We call this completed Euler trajectory the reference trajectory and denote its trace by $\mathcal{T}_a = (\{t_k\}_{k=0}^{a}, \{\bar{x}_k\}_{k=0}^{a}, \{\bar{v}_k\}_{k=0}^{a-1})$. Barred quantities remain fixed during correction. We write x_k and v_k for the state and velocity estimate along the corrected trajectory. Given $\mathcal{T}_a$ and access to f, FixUp-FM uses at most b new evaluations to produce the corrected endpoint x_a.

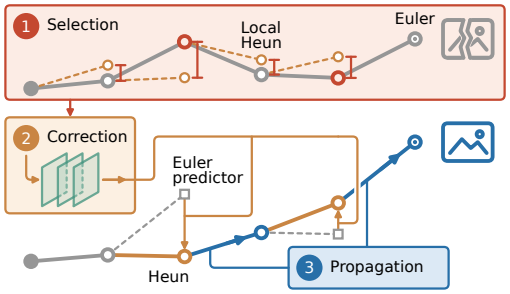


Figure 1: **FixUp-FM at a glance. 1** Select high-error intervals from the Euler trace. **2** Upgrade them to Heun. **3** Propagate the resulting state changes downstream.

3.1 Defect-based interval selection

With a limited budget, we first identify intervals with the largest local Euler errors from the completed trace. Euler advances using only the start velocity, so changes in velocity over the interval are not reflected in the update. Meanwhile, the endpoint velocity reveals this variation, providing a signal of the local error missed by Euler. For a nonterminal interval, both velocities are in the reference trace, so the corresponding Heun update is available. Its difference from Euler estimates the local Euler defect. We define this trace-based difference as the *reference correction*, with the corresponding interval score

$$\widetilde{d}_k := \frac{h_k}{2}(\bar{v}_{k+1} - \bar{v}_k), \qquad s_k := \|\widetilde{d}_k\|_2, \qquad 0 \le k \le a-2. \tag{2}$$

For the final interval, whose endpoint velocity is not evaluated by Euler, we use a one-sided estimate, setting $\widetilde{d}_{a-1} := \frac{h_{a-1}^2}{2h_{a-2}}(\bar{v}_{a-1} - \bar{v}_{a-2})$ and $s_{a-1} := \|\widetilde{d}_{a-1}\|_2$. FixUp-FM selects the K highest-scoring intervals. Let $S = \{p_1 < \cdots < p_K\}$ denote their indices in chronological order.

To justify this selection, let $\Phi_{t,h}(x)$ be the exact ODE flow from state x at time t over duration h, and define the Euler local defect $\tau_k^E(x) := \Phi_{t_k,h_k}(x) - [x + h_k f(x, t_k)]$. Write $\mathcal{R}_3 := \sum_{k=0}^{a-1} h_k^3 + h_{a-1}^2 h_{a-2} = O(h_{\max}^2)$ for the grid remainder, where $h_{\max} := \max_k h_k$.

Theorem 1 (Defect consistency and near-optimal allocation)**.** Assume $f \in C^2$ on a common compact neighborhood of the reference trajectory. With grid-independent constants, the reference correction estimates the Euler local defect with

$$\|\tau_k^E(\bar{x}_k) - \widetilde{d}_k\| \le C_d h_k^3 + \mathbf{1}_{\{k=a-1\}} C_\partial h_{a-1}^2 h_{a-2}. \tag{3}$$

Moreover, the selected set satisfies

$$\sum_{k \notin S} \|\tau_k^E(\bar{x}_k)\| \le \min_{|A|=K} \sum_{k \notin A} \|\tau_k^E(\bar{x}_k)\| + C_{\text{alloc}} \min\{\mathcal{R}_3, K h_{\max}^3\}. \tag{4}$$

Hence, the selection leaves at most $O(h_{\max}^2)$ more total reference Euler defect than the oracle K-interval selection, improving to $O(h_{\max}^3)$ for fixed K. If, in addition, $K < a$ and the ordered scores satisfy $s_{(K)} - s_{(K+1)} > 2\max_k |\|\tau_k^E(\bar{x}_k)\| - s_k|$, then S is exactly the top-K set of reference Euler defect norms (Appendix B).

3.2 Selective Heun correction and exact endpoint reuse

We now specify how each correction is performed. We upgrade the selected Euler steps to Heun, processing them in time order so that each correction starts from the state produced by earlier updates. Specifically, when FixUp-FM reaches a selected interval k with state x_k and velocity estimate v_k, it forms the predictor, obtains the endpoint stage, and applies Heun:

$$y_{k+1} = x_k + h_k v_k, \qquad w_{k+1} = f(y_{k+1}, t_{k+1}), \qquad x_{k+1} = x_k + \frac{h_k}{2}(v_k + w_{k+1}). \tag{5}$$

Relative to the Euler predictor, the *actual state correction* is $u_{k+1} := x_{k+1} - y_{k+1} = \frac{h_k}{2}(w_{k+1} - v_k)$. For a nonterminal interval starting from its reference state, this update gives the Heun endpoint $\bar{x}_{k+1}^H = \bar{x}_{k+1} + \widetilde{d}_k$, with $O(h_k^3)$ local error by equation 3.

At the earliest selected interval $p_1 < a - 1$, the predictor coincides with the reference endpoint, $y_{p_1+1} = \bar{x}_{p_1+1}$. The endpoint stage can therefore be reused exactly, giving $w_{p_1+1} = \bar{v}_{p_1+1}$ and $u_{p_1+1} = \widetilde{d}_{p_1}$ without a new evaluation. This endpoint-stage reuse is unique to Heun among explicit two-stage, second-order Runge–Kutta methods (Appendix C.1).

3.3 Minimum-norm directional propagation

However, once a correction moves the trajectory, the available velocities no longer correspond to the new downstream states. Re-evaluating every downstream velocity would undermine the fixed-budget objective, while directly reusing them would create a downstream state–velocity mismatch. FixUp-FM instead uses the local state–velocity relation encoded by the reference trace to adjust downstream velocities as the trajectory moves, allowing the correction to propagate without full downstream re-evaluation.

Let $\delta_j := x_j - \bar{x}_j$ denote the displacement from the reference trajectory and let $J_j := J_x f(\bar{x}_j, t_j)$ be the spatial Jacobian. For a C^2 field, a first-order spatial expansion at fixed time gives $f(x_j, t_j) = \bar{v}_j + J_j\delta_j + O(\|\delta_j\|^2)$. Thus, we only need the spatial response $J_j\delta_j$ to the correction displacement, rather than the full Jacobian. Adjacent velocities on the reference trajectory are evaluated at different states and times, so their difference does not isolate this response. We instead use three reference points to cancel the affine time dependence.

For each interior velocity node $1 \le j \le a - 2$, define the state and velocity chord residuals

$$\theta_j := \frac{h_{j-1}}{h_{j-1} + h_j}, \qquad \begin{aligned} q_j &:= \bar{x}_j - [(1-\theta_j)\bar{x}_{j-1} + \theta_j \bar{x}_{j+1}], \\ r_j &:= \bar{v}_j - [(1-\theta_j)\bar{v}_{j-1} + \theta_j \bar{v}_{j+1}]. \end{aligned} \tag{6}$$

They measure deviations from the respective time chords. Under the local affine model $f(x,t) = \bar{v}_j + A_j(x - \bar{x}_j) + g_j(t - t_j)$, the constant and time-dependent terms cancel, leaving $r_j = A_j q_j$, which is the complete spatial constraint imposed by the triplet. It identifies the response along q_j without determining the full spatial matrix (Appendix D.1).

Crucially, the reference Euler relations also give

$$q_j = -\frac{2h_j}{h_{j-1} + h_j}\widetilde{d}_{j-1}. \tag{7}$$

Thus, **the direction identified by the triplet is precisely the one needed to adjust the velocity** after a Heun correction from the reference state. As corrections accumulate, however, the current displacement δ_j may no longer lie in $\mathrm{span}(q_j)$. For $q_j \neq 0$, FixUp-FM extends the observed response with the minimum-Frobenius-norm map satisfying $Bq_j = r_j$:

$$B_j^\star = \frac{r_j q_j^\top}{\|q_j\|^2}, \qquad v_j = \bar{v}_j + B_j^\star \delta_j = \bar{v}_j + r_j \frac{q_j^\top \delta_j}{\|q_j\|^2}. \tag{8}$$

We set $B_j^\star = 0$ at zero chords $q_j = 0$ and at boundary velocity nodes $j \in \{0, a-1\}$, where the triplet is unavailable. Each $B_j^\star$ depends only on the reference trace and can be precomputed once.

Theorem 2 (Minimum-norm extension and exact action recovery)**.** For $q_j \neq 0$, the minimizer $B_j^\star$ is unique. It also attains the minimum operator norm $\|B_j^\star\|_2 = \|r_j\|/\|q_j\|$ and satisfies $B_j^\star z = 0$ for $z \perp q_j$, minimizing worst-case amplification without responding to orthogonal displacements.

At an interior velocity node j, a correction on interval $j - 1$ from its reference state gives $\delta_j = \widetilde{d}_{j-1} \in \mathrm{span}(q_j)$. Under the local affine model over the reference triplet and corrected state,

$$B_j^\star \delta_j = A_j \delta_j, \qquad v_j = f(\bar{x}_j + \delta_j, t_j). \tag{9}$$

When $q_j = 0$, the same conclusion holds since $\widetilde{d}_{j-1} = \delta_j = 0$. Thus, this correction requires no extrapolation beyond the trace-identified direction, and its propagated velocity is exact in the affine setting (Appendix D).

3.4 Complete algorithm and anytime continuation

Together, starting from p_1, FixUp-FM applies Heun on selected intervals and Euler elsewhere, propagating the velocity after each update. The displacement then evolves as

$$\delta_{k+1} = (I + h_k B_k^\star)\delta_k + \mathbf{1}_{\{k \in S\}} u_{k+1}. \tag{10}$$

Algorithm 1 FixUp-FM

Require: velocity field f; completed Euler trace $\mathcal{T}_a$; correction count $K \in \{1, \dots, a\}$

Compute scores as in Section 3.1; select the top-K indices $S = \{p_1 < \cdots < p_K\}$
Compute $B_j^\star$ by equation 8, with the zero and boundary fallbacks
Initialize $x_{p_1} \leftarrow \bar{x}_{p_1}$, $v_{p_1} \leftarrow \bar{v}_{p_1}$
for $k = p_1, \dots, a-1$ **do**
    if $k \in S$ **then**
        Apply equation 5, reusing the endpoint velocity $w_{k+1} = \bar{v}_{k+1}$ if $k = p_1 < a-1$
    else
        $x_{k+1} \leftarrow x_k + h_k v_k$
    end if
    if $k < a-1$ **then**
        $v_{k+1} \leftarrow \bar{v}_{k+1} + B_{k+1}^\star(x_{k+1} - \bar{x}_{k+1})$
    end if
end for
return x_a

As summarized in Algorithm 1 and Figure 2, K corrections use $K-1$ new NFEs unless $S = \{a-1\}$, which requires one. We choose K so that this cost does not exceed b. This count is worst-case optimal for exact endpoint-stage evaluation when the remaining targets are distinct and not previously evaluated (Appendix C). For latent dimension d, non-network work costs $O(ad + a \log K)$, with heap-based selection and rank-one propagation, and trace storage costs $O(ad)$.

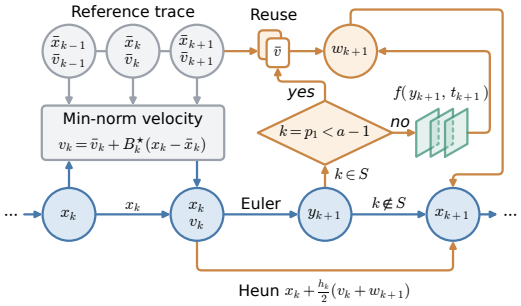


Figure 2: **A single FixUp-FM pass.** Interval selection, selective Heun correction, and trace-derived directional propagation.

Anytime continuation. FixUp-FM can be applied repeatedly. Each pass applies the same correction and propagation procedure to the current trajectory, re-scoring and selecting only uncorrected intervals using the current velocities. The reference trace and $B_k^\star$ remain fixed across passes. Let c_k denote the correction u_{k+1} first applied on interval k, with $c_k = 0$ until that interval is corrected. Each completed pass then satisfies

$$x_{k+1} = x_k + h_k v_k + c_k, \tag{11}$$

where v_k follows equation 8. Since later passes start from an already corrected trajectory, they do not generally retain the initial endpoint-stage reuse. Any pass after the first therefore uses at most K_{req} new evaluations for K_{req} newly selected intervals. Under the fixed-budget assumptions, the refreshed scores still approximate the current Euler defect norms to third order at nonterminal intervals. The terminal score has the same rate when h_{a-1}/h_{a-2} remains uniformly bounded. Under the same condition, selecting a fixed total of K intervals across requests preserves the near-oracle guarantee of Theorem 1 on the final corrected trajectory (Appendix H).

4 Numerical Analysis

We now study how FixUp-FM's local design choices translate into global accuracy. Since the propagation rule relies on a first-order local approximation, we assume $f \in C^2$ on a common compact neighborhood containing the reference and corrected states, stage points, and exact or ideal-rerun comparison trajectories. This provides the local smoothness and uniform bounds on the relevant derivatives used below. We use Euclidean norms for vectors and their induced operator norms for matrices, and take all constants to be independent of the time grid.

4.1 ACCURACY AND STABILITY OF DIRECTIONAL PROPAGATION

We begin with the affine setting, where the exact action in Theorem 2 can persist downstream. Suppose $f(x,t) = Ax + gt + \beta$ holds over a correction-free uniform-grid segment, including the corresponding reference triplets and corrected states. Starting from an aligned node s with $\delta_s \in \mathrm{span}(q_s)$, the displacement remains aligned downstream, yielding $v_j = f(x_j, t_j)$ at every interior velocity node. Thus, the resulting Euler updates exactly match full velocity re-evaluation.

For a general nonlinear field, define the chord non-affinity residual $\varepsilon_j := r_j - J_j q_j$. For $q_j \neq 0$, let $P_j := q_j q_j^\top / \|q_j\|^2$ denote the orthogonal projector onto $\mathrm{span}(q_j)$. We measure the propagation error by $\eta_j := \|f(x_j, t_j) - v_j\|$.

Proposition 1 (Nonlinear velocity-extension error)**.** Let M bound $\|D^2 f\|$ on the common neighborhood. At an interior node with $q_j \neq 0$,

$$\eta_j \leq \|J_j\| \, \|(I - P_j)\delta_j\| + \frac{\|\varepsilon_j\|}{\|q_j\|} \, \|P_j \delta_j\| + \frac{M}{2} \|\delta_j\|^2, \tag{12}$$

where $\|\varepsilon_j\| = O(h_{j-1} h_j)$. The three terms respectively capture off-direction displacement, local non-affinity, and the spatial Taylor remainder. For a correction from a reference state, Theorem 2 gives $\delta_j = \widetilde{d}_{j-1} \in \mathrm{span}(q_j)$, eliminating the off-direction term. Together, these estimates yield $\eta_j = O(h_{j-1}(h_{j-1} + h_j)) = O(h_{\max}^2)$, including on nonuniform grids. The zero-chord and boundary cases also satisfy the $O(h_{\max}^2)$ bound.

Displacement stability. To control the downstream displacement appearing in this error bound, define the directional amplification $\gamma_j := \|B_j^\star \delta_j\| / \|\delta_j\|$, with $\gamma_j = 0$ when $\delta_j = 0$. For $s < m$, the recursion equation 10 gives

$$\|\delta_m\| \leq \exp\left(\sum_{j=s}^{m-1} h_j \gamma_j \right) \left[\|\delta_s\| + \sum_{\ell=s+1}^{m} \mathbf{1}_{\{\ell-1 \in S\}} \|u_\ell\| \right]. \tag{13}$$

Thus, downstream displacement growth is controlled by the initial displacement, realized directional amplification, and correction injections, with exponential factor at most $e^{\max_j \|B_j^\star\|_2}$ (Appendix E).

4.2 END-TO-END ACCURACY AND FIXED-BUDGET CONSISTENCY

We now connect the local propagation bounds to end-to-end accuracy. For a nonterminal interval, the reference correction differs from the current Euler defect by $\|\tau_k^E(x_k) - \widetilde{d}_k\| = O(h_k^3 + h_k^2 \|\delta_k\|)$. At the terminal interval, the one-sided estimator contributes an additional $O(h_{a-1}^2 h_{a-2})$ term.

We next bound the local update error under the propagated velocity. Let L be a Lipschitz constant of f in the state variable. An unselected Euler step has local error at most $\|\tau_k^E(x_k)\| + h_k \eta_k$, while a selected Heun step removes the leading Euler defect, leaving at most $C_H h_k^3 + \frac{h_k}{2}(1 + L h_k)\eta_k$. Aggregating these contributions yields the following result.

Theorem 3 (End-to-end defect decomposition)**.** Let $x^\star(1)$ be the exact ODE endpoint from the same initial latent. Then

$$\|x^\star(1) - x_a\| \leq C \left[\sum_{k \notin S} \|\widetilde{d}_k\| + \mathcal{R}_3 + \sum_{k \notin S} h_k^2 \|\delta_k\| + \sum_{k=0}^{a-1} h_k \eta_k \right]. \tag{14}$$

The four terms represent the estimated defects on unselected intervals, higher-order remainders, changes in local defects as the states move, and velocity-extension error. Top-K selection exactly minimizes the first term, while $\mathcal{R}_3$ depends only on the time grid. To control the last two terms, suppose $\max_j \|B_j^\star\|_2 \leq M_B$ uniformly. Then both $\max_j \|\delta_j\|$ and $\max_j \eta_j$ are $O(\sum_{k \in S} h_k^2)$, uniformly in the selected set.

For each reference trace, the propagation norms in the assumption are directly computable as $\|B_j^\star\|_2 = \|r_j\| / \|q_j\|$ for nonzero interior chords.

Corollary 1 (Fixed-budget correction consistency)**.** Under the uniform propagation bound above, with K fixed as $h_{\max} \to 0$, the trajectory displacement and velocity-extension error satisfy

$$\max_j \|\delta_j\| = O(h_{\max}^2), \qquad \max_j \eta_j = O(h_{\max}^2). \tag{15}$$

These bounds preserve the third-order accuracy of the reference estimates as the trajectory changes. The actual corrections satisfy $\max_{k\in S} \|u_{k+1} - \widetilde{d}_k\| = O(h_{\max}^3)$. The reference-to-current estimate also gives $\max_k \|\tau_k^E(x_k) - \widetilde{d}_k\| = O(h_{\max}^3)$, so the reference scores approximate the current Euler defect norms to third order. Hence, for $K < a$, the exact top-K guarantee of Theorem 1 holds whenever $s_{(K)} - s_{(K+1)} > 2C_{\mathrm{cur}} h_{\max}^3$ for some constant C_{cur}. Otherwise, selection leaves at most $O(Kh_{\max}^3)$ more total Euler defect than the oracle selection for the corrected trajectory.

To isolate the effect of velocity approximation on the endpoint, let $x_j^{\mathrm{ideal}(S)}$ denote the state at t_j of the selective Euler–Heun rerun from the same initial latent, on the same grid and selected set, with exact current velocities at every step. Under the same uniform propagation bound,

$$\max_{0\le j\le a} \|x_j - x_j^{\mathrm{ideal}(S)}\| = O\left(\sum_{k\in S} h_k^2\right), \tag{16}$$

uniformly in the selected set. For fixed K, this gap is $O(h_{\max}^2)$, so **trace-derived propagation does not alter the ideal selective rerun's leading first-order behavior**. Under affine dynamics on a uniform grid, this sharpens to $O(Kh_{\max}^3 + K^2 h_{\max}^4)$.

Even when all intervals are selected, if the propagation error remains $\max_k \eta_k = O(h_{\max}^2)$, equation 14 gives $\|x^\star(1) - x_a\| = O(h_{\max}^2)$, so propagation preserves Heun's second-order convergence in this limiting case. Proofs of these one-shot results are in Appendix F. Finally, we compare successive requests with a single pass at the same total correction count.

Proposition 2 (Consistency of anytime continuation)**.** Let x_j^{any} and x_j^{one} denote the states at t_j of continuation and one-shot runs from the same trace, each correcting a fixed total of K intervals. Assume the conditions of Corollary 1 hold uniformly for both runs. Then

$$\max_{0\le j\le a} \|x_j^{\mathrm{any}} - x_j^{\mathrm{one}}\| = \begin{cases} O(h_{\max}^3), & \text{if their selected sets coincide,} \\ O(h_{\max}^2), & \text{otherwise,} \end{cases} \tag{17}$$

uniformly over request groupings and correction orderings. Thus, splitting a fixed set of corrections across requests introduces only a third-order endpoint perturbation, while re-scoring that changes the selected set still preserves second-order agreement with a one-shot correction (Appendix G).

5 Experiment

5.1 Experiment setup

We evaluate SD3 Medium (Esser et al., 2024), SD3.5 Large (Stability AI, 2024), and FLUX.1-dev (Black Forest Labs, 2024) on FluxPrompting prompts (Gourbin, 2024), using 512×512 as the default resolution. Euler references use each model's native timestep schedule; network evaluations use FP16 and guidance scale 3.5. We use ImageReward (IR) (Xu et al., 2023a) for preference-related image quality and CLIP similarity (Radford et al., 2021) for text–image alignment, computed with PE-Core-bigG-14-448 (Bolya et al., 2025). Within each experiment, methods share prompts and initial noise; post-sampling methods additionally start from the same completed Euler output.

For the main comparison, we use 5,000 prompts per model and evaluate quality–cost trade-offs over initial budgets $a = 8, 12, \ldots, 32$. For FixUp-FM, we sweep correction counts K independently at each a, with all runs starting from the same reference trace. Displayed correction counts are selected on 1,024 independent validation prompts per model. We compare against direct Euler, DPM-adaptive (Lu et al., 2022; Crowson, 2022), DOPRI5 (Dormand & Prince, 1980), and SRDS (Selvam et al., 2024) as sampling baselines, and Restart (Xu et al., 2023b) and SDEdit (Meng et al., 2022) as post-sampling methods. We sweep error tolerances for the adaptive solvers and coarse–fine grids for SRDS; Restart and SDEdit each use one noise strength per model, jointly tuned across seven

Table 1: **Post-sampling comparisons.** Mean ImageReward (IR) and CLIP on 5,000 prompts; ↑ indicates higher is better. a is the initial Euler budget. Bold marks the highest IR for each model and base budget. Superscripts give mean added NFE; $(+0)$ means no new evaluations. Euler matches FixUp-FM's samplewise total NFE.

a	Metric	SD3 Medium				SD3.5 Large				FLUX.1-dev			
		Euler	FixUp	SDEdit	Restart	Euler	FixUp	SDEdit	Restart	Euler	FixUp	SDEdit	Restart
8	IR↑	0.3831	**0.4541**$^{(+0)}$	0.4026$^{(+1)}$	0.3459$^{(+1)}$	0.2741	**0.3410**$^{(+0)}$	0.2905$^{(+1)}$	0.2229$^{(+1)}$	**0.7761**	0.7747$^{(+1)}$	0.7496$^{(+1)}$	0.7212$^{(+1)}$
	CLIP↑	26.23	26.18	26.13	26.20	26.40	26.36	26.35	26.36	26.47	26.60	26.59	26.40
12	IR↑	0.5877	**0.6377**$^{(+0)}$	0.6063$^{(+1)}$	0.5690$^{(+1)}$	0.4708	**0.5285**$^{(+0)}$	0.4869$^{(+1)}$	0.4390$^{(+1)}$	0.8383	**0.8543**$^{(+2)}$	0.8475$^{(+2)}$	0.8236$^{(+3)}$
	CLIP↑	26.50	26.48	26.45	26.48	26.70	26.68	26.67	26.67	26.47	26.61	26.62	26.70
16	IR↑	0.6434	**0.6767**$^{(+0)}$	0.6589$^{(+1)}$	0.6285$^{(+1)}$	0.5365	**0.5779**$^{(+0)}$	0.5500$^{(+1)}$	0.5111$^{(+1)}$	0.8603	0.8734$^{(+2)}$	**0.8756**$^{(+2)}$	0.8563$^{(+3)}$
	CLIP↑	26.53	26.51	26.51	26.52	26.78	26.77	26.78	26.77	26.47	26.56	26.64	26.73
20	IR↑	0.6885	**0.7110**$^{(+1)}$	0.6874$^{(+1)}$	0.6618$^{(+1)}$	0.5670	**0.5997**$^{(+1)}$	0.5733$^{(+1)}$	0.5389$^{(+1)}$	0.8695	0.8846$^{(+3)}$	**0.8877**$^{(+3)}$	0.8734$^{(+3)}$
	CLIP↑	26.60	26.59	26.58	26.57	26.80	26.80	26.81	26.78	26.43	26.49	26.55	26.71
24	IR↑	0.7149	**0.7361**$^{(+3)}$	0.7175$^{(+3)}$	0.7044$^{(+3)}$	0.5952	**0.6193**$^{(+3)}$	0.5939$^{(+3)}$	0.5871$^{(+3)}$	0.8804	0.8959$^{(+5)}$	0.8931$^{(+5)}$	**0.9085**$^{(+5)}$
	CLIP↑	26.63	26.65	26.62	26.60	26.80	26.82	26.82	26.81	26.40	26.45	26.47	26.68
28	IR↑	0.7223	**0.7455**$^{(+1)}$	0.7303$^{(+1)}$	0.7107$^{(+1)}$	0.6105	**0.6333**$^{(+6)}$	0.6056$^{(+6)}$	0.6041$^{(+7)}$	0.8799	0.8990$^{(+7)}$	0.8925$^{(+7)}$	**0.9128**$^{(+7)}$
	CLIP↑	26.65	26.65	26.65	26.63	26.82	26.82	26.80	26.80	26.35	26.42	26.43	26.55
32	IR↑	0.7389	**0.7559**$^{(+1)}$	0.7429$^{(+1)}$	0.7258$^{(+1)}$	0.6153	**0.6405**$^{(+7)}$	0.6133$^{(+7)}$	0.6123$^{(+7)}$	0.8789	0.8928$^{(+7)}$	0.8888$^{(+7)}$	**0.9103**$^{(+7)}$
	CLIP↑	26.68	26.68	26.67	26.65	26.82	26.82	26.81	26.81	26.34	26.40	26.41	26.54

calibration base/budget pairs on the same 64 held-out prompts. Direct Euler is matched samplewise to FixUp-FM's actual total NFE, while SDEdit and Restart match its added budget when possible and otherwise use the next higher native budget.

For anytime continuation, we use 3,000 prompts per model from Euler-32 references using Section 3.4. FixUp-FM and SDEdit are evaluated with $+1 \times 6$, $+2 \times 3$, $+3 \times 2$, and a single $+6$ request, while Restart uses $+1 \times 6$, $+3 \times 2$, and a single $+5$ request, following its native odd-valued budgets. All schedules start from the same initial output, and repeated requests continue from the latest output. Restart and SDEdit use the same parameters as in the main comparison. We score every completed request and report quality against mean cumulative added NFE.

For ablations, we use SD3.5 Large and 1,000 prompts disjoint from the configuration-screening set. We compare defect-based selection with uniform and random selection at $a = 24$, $K = 2, 4, 6$, with correction and propagation rules unchanged. At $a = 24, 28, 32$ with $K = 5$, we compare minimum-norm propagation with unchanged reference velocities while holding the selected intervals fixed. In addition, we evaluate FixUp-FM against NFE-matched Euler at 1024×1024 from Euler-24 references with $K = 0, \ldots, 8$, and measure 512×512 end-to-end latency on an A100 at batch size one. Details are provided in Appendices I and J.

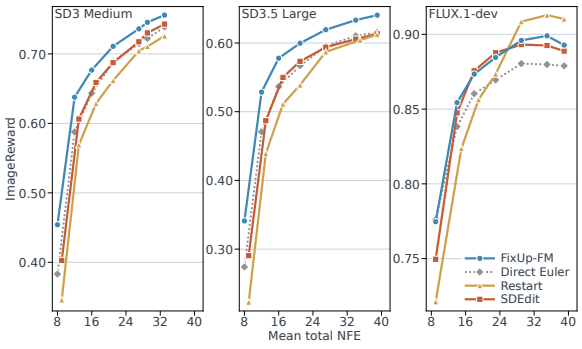


Figure 3: **Image quality versus total NFE.** Mean ImageReward on 5,000 FluxPrompting prompts per model at 512×512. Refinement starts from shared Euler outputs with $a = 8, 12, \ldots, 32$. Direct Euler is NFE-matched to FixUp-FM per sample. Table 1 gives budgets and CLIP scores.

5.2 Main results

Across all tested base budgets, FixUp-FM improves IR over the initial Euler output on all three models. At the same total NFE, it achieves higher IR than direct Euler at 20 of the 21 points in Figure 3 and Table 1 and closely matches it at the remaining one. This advantage extends to 111 of 120 configurations across the full K sweeps. The gains reach 0.0710, 0.0669, and 0.0196 on SD3, SD3.5, and FLUX, respectively. At $a = 12$, the displayed configurations improve IR over NFE-matched Euler by 0.0500, 0.0577, and 0.0160, with 0, 0, and 2 additional NFEs, respectively.

Against the post-sampling baselines, FixUp-FM achieves higher IR than both SDEdit and Restart at all displayed points on SD3 and SD3.5. On FLUX, it exceeds SDEdit at 5 of 7 points and closely matches it at the remaining two, while outperforming Restart at the four lower-budget points. CLIP scores remain broadly similar across these methods.

Compared with sampling baselines, FixUp-FM achieves higher IR and CLIP with fewer NFEs than DPM-adaptive and DOPRI5 across all three models. It also generally outperforms SRDS, with both IR and CLIP consistently higher on SD3 and SD3.5, while on FLUX it performs better at low NFEs and remains competitive in IR at intermediate budgets.

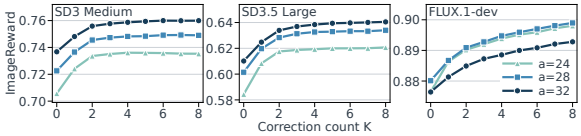


Figure 4: **Effect of the correction budget.** Mean ImageReward of FixUp-FM on 5,000 prompts per model at 512×512, for $a = 24, 28, 32$ and $K = 0, \ldots, 8$. $K = 0$ is uncorrected Euler.

The full K sweeps show a stable correction pattern, with gains saturating after the first few corrections on SD3 and SD3.5 and accumulating gradually on FLUX. The overall trend holds at 1024×1024. At 512×512, FixUp-FM adds less than 0.7% end-to-end latency over NFE-matched Euler across all timed configurations. Full results are provided in Appendices K, O, and P.

5.3 Anytime Continuation

Figure 5 shows that FixUp-FM sustains its quality gains under successive requests. Across all three models, mean IR remains above both the initial Euler output and NFE-matched direct Euler throughout the evaluated continuation schedules. On SD3 and SD3.5, distributing the same correction budget over multiple requests yields IR and CLIP close to a one-shot correction. Across the evaluated budgets, the maximum IR differences across request groupings are only 0.0020 and 0.0046 on SD3 and SD3.5, respectively.

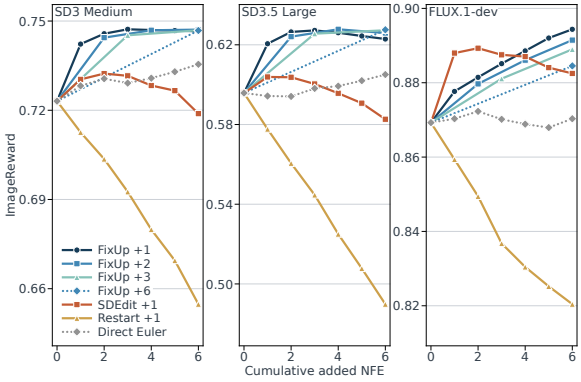


Figure 5: **Anytime continuation.** Mean ImageReward on 3,000 prompts per model from Euler-32 outputs. FixUp-FM uses one-, two-, or three-evaluation requests, or a single six-evaluation request; SDEdit and Restart use one evaluation per request. Direct Euler is sampled afresh at the corresponding total NFE.

On FLUX, all continuation schedules improve monotonically in both IR and CLIP, with finer-grained requests performing better at the same cumulative budget. Six one-evaluation requests exceed a single six-evaluation request by 0.0098 IR, indicating an additional benefit from incremental refinement.

This stability contrasts with repeated post-sampling refinement. Under repeated one-evaluation requests, SDEdit peaks early and then declines, while Restart degrades steadily on all three models. Across the tested request groupings, FixUp-FM stays above both baselines throughout on SD3 and SD3.5. On FLUX, its $+1$ continuation overtakes SDEdit after four added NFEs and reaches 0.8944 IR at $+6$, compared with 0.8825 for SDEdit and 0.8204 for Restart under the same request schedule. Full results are provided in Appendix M.

5.4 Ablation

Both defect-based selection and minimum-norm propagation contribute consistent matched-NFE gains. Defect-based selection improves IR over uniform and random selection by 0.0084–0.0241 at matched NFE across all tested budgets. With the selected intervals fixed, minimum-norm propagation improves IR by 0.0055–0.0080 over unchanged reference velocities at the same NFE. All paired 95% confidence intervals for these IR improvements are above zero (Appendix N).

AI USE STATEMENT

We used generative AI tools to assist with manuscript drafting and revision, literature search, parts of the theoretical derivations and proofs, and the design and implementation of the affine numerical examples. The authors reviewed and revised all AI-assisted text and checked the cited references against their original sources. The authors take full responsibility for the final content of this work, including all claims, results, and citations.

REFERENCES

Heli Ben-Hamu, Omri Puny, Itai Gat, Brian Karrer, Uriel Singer, and Yaron Lipman. D-Flow: Differentiating through flows for controlled generation. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 3462–3483. PMLR, 2024. URL https://proceedings.mlr.press/v235/ben-hamu24a.html.

Black Forest Labs. FLUX.1 [dev]. Hugging Face model card, 2024. URL https://huggingface.co/black-forest-labs/FLUX.1-dev.

Daniel Bolya, Po-Yao Huang, Peize Sun, Jang Hyun Cho, Andrea Madotto, Chen Wei, Tengyu Ma, Jiale Zhi, Jathushan Rajasegaran, Hanoona Abdul Rasheed, Junke Wang, Marco Monteiro, Hu Xu, Shiyu Dong, Nikhila Ravi, Shang-Wen Li, Piotr Dollár, and Christoph Feichtenhofer. Perception encoder: The best visual embeddings are not at the output of the network. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL https://papers.neurips.cc/paper_files/paper/2025/hash/57bc0a850255e2041341bf74c7e2b9fa-Abstract-Conference.html.

Katherine Crowson. k-diffusion. Software repository, 2022. URL https://github.com/crowsonkb/k-diffusion.

J. E. Dennis, Jr. and R. B. Schnabel. Least change secant updates for quasi-Newton methods. *SIAM Review*, 21(4):443–459, 1979. doi: 10.1137/1021091.

J. R. Dormand and P. J. Prince. A family of embedded runge–kutta formulae. *Journal of Computational and Applied Mathematics*, 6(1):19–26, 1980. doi: 10.1016/0771-050X(80)90013-3. URL https://www.sciencedirect.com/science/article/pii/0771050X80900133.

Noam Elata, Bahjat Kawar, Tomer Michaeli, and Michael Elad. Nested diffusion processes for anytime image generation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 5018–5027, 2024. URL https://openaccess.thecvf.com/content/WACV2024/html/Elata_Nested_Diffusion_Processes_for_Anytime_Image_Generation_WACV_2024_paper.html.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, Dustin Podell, Tim Dockhorn, Zion English, and Robin Rombach. Scaling rectified flow transformers for high-resolution image synthesis. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 12606–12633. PMLR, 2024. URL https://proceedings.mlr.press/v235/esser24a.html.

Luca Eyring, Shyamgopal Karthik, Karsten Roth, Alexey Dosovitskiy, and Zeynep Akata. ReNO: Enhancing one-step text-to-image models through reward-based noise optimization. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/e31bdea0a93741c2157eea705dd219eb-Abstract-Conference.html.

Vincent Gourbin. FluxPrompting. Hugging Face dataset, 2024. URL https://huggingface.co/datasets/VincentGOURBIN/FluxPrompting.

Ernst Hairer, Syvert P. Nørsett, and Gerhard Wanner. *Solving Ordinary Differential Equations I: Nonstiff Problems*, volume 8 of *Springer Series in Computational Mathematics*. Springer, Berlin, Heidelberg, 2 edition, 1993. doi: 10.1007/978-3-540-78862-1.

Jaihoon Kim, Taehoon Yoon, Jisung Hwang, and Minhyuk Sung. Inference-time scaling for flow models via stochastic generation and rollover budget forcing. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/2c71b14637802ed08eaa3cf50342b2b9-Abstract-Conference.html.

Inho Kong, Sojin Lee, Youngjoon Hong, and Hyunwoo J. Kim. Error as signal: Stiffness-aware diffusion sampling via embedded runge-kutta guidance. In *International Conference on Learning Representations*, 2026.

Gyubin Lee, Bao Truong, Jaesik Yoon, Dongwoo Lee, Minsu Kim, Yoshua Bengio, and Sungjin Ahn. Adaptive inference-time scaling via cyclic diffusion search. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

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, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/260a14acce2a89dad36adc8eefe7c59e-Abstract-Conference.html.

Nanye Ma, Shangyuan Tong, Haolin Jia, Hexiang Hu, Yu-Chuan Su, Mingda Zhang, Xuan Yang, Yandong Li, Tommi Jaakkola, Xuhui Jia, and Saining Xie. Scaling inference time compute for diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2523–2534, 2025.

Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. SDEdit: Guided image synthesis and editing with stochastic differential equations. In *International Conference on Learning Representations*, 2022. URL https://sde-image-editing.github.io/.

Xin Peng and Ang Gao. Rethinking refinement: Correcting generative bias without noise injection, 2026. URL https://arxiv.org/abs/2601.21182.

Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. SDXL: Improving latent diffusion models for high-resolution image synthesis. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/081b08068e4733ae3e7ad019fe8d172f-Abstract-Conference.html.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763. PMLR, 2021. URL https://proceedings.mlr.press/v139/radford21a.html.

Nikil Roashan Selvam, Amil Merchant, and Stefano Ermon. Self-refining diffusion samplers: Enabling parallelization via parareal iterations. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/09d320a5b92d74bbde3d0c4f52e680a9-Abstract-Conference.html.

Stability AI. Stable diffusion 3.5 large. Hugging Face model card, 2024. URL https://huggingface.co/stabilityai/stable-diffusion-3.5-large.

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, 2023a.

Yilun Xu, Mingyang Deng, Xiang Cheng, Yonglong Tian, Ziming Liu, and Tommi Jaakkola. Restart sampling for improving generative processes. In *Advances in Neural Information Processing Systems*, volume 36, 2023b. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/f2543511e5f4d4764857f9ad833a977d-Abstract-Conference.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, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/9c2aa1e456ea543997f6927295196381-Abstract-Conference.html.

Wenliang Zhao, Haolin Wang, Jie Zhou, and Jiwen Lu. DC-Solver: Improving predictor-corrector diffusion sampler via dynamic compensation. In *European Conference on Computer Vision*, 2024. URL https://www.ecva.net/papers/eccv_2024/papers_ECCV/html/1795_ECCV_2024_paper.php.

A Assumptions and notation

Let $\Phi_{t,h}(x)$ denote the exact solution at time $t+h$ of

$$\dot{z}(s) = f(z(s), s), \qquad z(t) = x. \tag{A.1}$$

We use the time grid and trace from Section 3, with $a \geq 2$. Unless a result states weaker assumptions, all reference and corrected states, stage points, and the exact and ideal-rerun states used in a comparison lie in a common compact neighborhood $\Omega \subset \mathbb{R}^d$. The field is C^2 on an open neighborhood of this set and the relevant times; in particular,

$$f \in C^2(\Omega \times [0,1]). \tag{A.2}$$

Consequently, f, $J_x f$, $\partial_t f$, and the second derivatives of f are uniformly bounded on the relevant neighborhood. The material derivative

$$\mathcal{M}(x,t) := J_x f(x,t) f(x,t) + \partial_t f(x,t) \tag{A.3}$$

has a bounded derivative on this neighborhood. Its Lipschitz bound applies along the state–time segments used in the Taylor expansions below. All constants may depend on these neighborhood bounds and the unit integration horizon, but not on the grid size. Vector norms are Euclidean; matrix operator norms are induced Euclidean norms; and $\|\cdot\|_F$ is the Frobenius norm. We write $\tau_k^E = \tau_k^E(\bar{x}_k)$ for the reference Euler defect. The local affine neighborhood in Theorem 2 is denoted by $\mathcal{N}_j$.

B Proof of Theorem 1

B.1 Nonterminal defect consistency and the reference Heun error

Fix $k \leq a-2$ and abbreviate

$$x = \bar{x}_k, \qquad t = t_k, \qquad h = h_k, \qquad v = f(x,t).$$

Along the exact solution issued from (x,t),

$$\ddot{x}(t) = J_x f(x(t),t)\dot{x}(t) + \partial_t f(x(t),t) = \mathcal{M}(x(t),t).$$

Taylor expansion of the exact flow (Hairer et al., 1993) gives

$$\Phi_{t,h}(x) = x + hv + \frac{h^2}{2}\mathcal{M}(x,t) + R_x, \qquad \|R_x\| \leq C_x h^3. \tag{B.1}$$

The Euler local defect is therefore

$$\tau_k^E = \frac{h^2}{2}\mathcal{M}(x,t) + R_x. \tag{B.2}$$

A first-order joint state–time expansion at the Euler endpoint gives

$$\begin{aligned} f(x+hv, t+h) &= f(x,t) + hJ_x f(x,t)v + h\partial_t f(x,t) + R_v \\ &= v + h\mathcal{M}(x,t) + R_v, \qquad \|R_v\| \leq C_v h^2. \end{aligned} \tag{B.3}$$

Because $\bar{x}_{k+1} = x + hv$, the recorded velocity is $\bar{v}_{k+1} = f(x+hv, t+h)$, and hence

$$\widetilde{d}_k = \frac{h}{2}(\bar{v}_{k+1} - \bar{v}_k) = \frac{h^2}{2}\mathcal{M}(x,t) + \frac{h}{2}R_v. \tag{B.4}$$

Subtracting equation B.2 and equation B.4 yields

$$\|\tau_k^E - \widetilde{d}_k\| \leq \left(C_x + \frac{C_v}{2}\right) h_k^3. \tag{B.5}$$

The reference Heun endpoint is $\bar{x}_{k+1}^H = \bar{x}_{k+1} + \widetilde{d}_k$. Therefore

$$\Phi_{t_k,h_k}(\bar{x}_k) - \bar{x}_{k+1}^H = \tau_k^E - \widetilde{d}_k,$$

and equation B.5 also proves the $O(h_k^3)$ local Heun error stated in Section 3.2.

B.2 Final one-sided estimator

Applying equation B.3 to the preceding reference interval gives

$$\bar{v}_{a-1} - \bar{v}_{a-2} = h_{a-2}\mathcal{M}(\bar{x}_{a-2}, t_{a-2}) + R^v_{a-2}, \qquad \|R^v_{a-2}\| \le Ch^2_{a-2}. \tag{B.6}$$

Thus

$$\frac{\bar{v}_{a-1} - \bar{v}_{a-2}}{h_{a-2}} = \mathcal{M}(\bar{x}_{a-2}, t_{a-2}) + O(h_{a-2}). \tag{B.7}$$

Since $\mathcal{M}$ is Lipschitz and

$$\|\bar{x}_{a-1} - \bar{x}_{a-2}\| = h_{a-2}\|\bar{v}_{a-2}\| = O(h_{a-2}),$$

we also have

$$\mathcal{M}(\bar{x}_{a-2}, t_{a-2}) = \mathcal{M}(\bar{x}_{a-1}, t_{a-1}) + O(h_{a-2}). \tag{B.8}$$

Multiplying equation B.7–equation B.8 by $h^2_{a-1}/2$ gives

$$\widetilde{d}_{a-1} = \frac{h^2_{a-1}}{2}\mathcal{M}(\bar{x}_{a-1}, t_{a-1}) + O(h^2_{a-1}h_{a-2}). \tag{B.9}$$

Equation B.2, applied at the final interval, gives

$$\tau^E_{a-1} = \frac{h^2_{a-1}}{2}\mathcal{M}(\bar{x}_{a-1}, t_{a-1}) + O(h^3_{a-1}). \tag{B.10}$$

Subtracting proves

$$\|\tau^E_{a-1} - \widetilde{d}_{a-1}\| \le C_\partial\big(h^3_{a-1} + h^2_{a-1}h_{a-2}\big). \tag{B.11}$$

Together with equation B.5, this proves the estimator bound in Theorem 1.

B.3 Additive near-oracle allocation

Write $d_k = \|\tau^E_k\|$, $\epsilon_k = |d_k - s_k|$, and $\epsilon_{\max} = \max_k \epsilon_k$. The reverse triangle inequality and the estimator bounds give

$$\sum_{k=0}^{a-1} \epsilon_k \le C\mathcal{R}_3, \qquad \epsilon_{\max} \le Ch^3_{\max}, \qquad \mathcal{R}_3 = \sum_{k=0}^{a-1} h^3_k + h^2_{a-1}h_{a-2}. \tag{B.12}$$

Let $A_\star$ minimize the true unresolved defect over sets of size K. Since S contains the K largest scores, $\sum_{k\in A_\star\setminus S} s_k \le \sum_{k\in S\setminus A_\star} s_k$. Thus

$$\begin{aligned}\sum_{k\notin S} d_k - \sum_{k\notin A_\star} d_k &= \sum_{k\in A_\star\setminus S} d_k - \sum_{k\in S\setminus A_\star} d_k \\ &\le \sum_{k\in A_\star\triangle S} \epsilon_k \le \min\left\{\sum_k \epsilon_k, 2\min\{K, a-K\}\epsilon_{\max}\right\}.\end{aligned} \tag{B.13}$$

Combining the two estimates yields

$$\sum_{k\notin S} d_k \le \min_{|A|=K} \sum_{k\notin A} d_k + C\min\{\mathcal{R}_3, Kh^3_{\max}\}. \tag{B.14}$$

The constant is independent of the grid and selected set. Since $\sum_k h_k = 1$, we have

$$\sum_k h^3_k \le h^2_{\max}, \qquad h^2_{a-1}h_{a-2} \le h^3_{\max},$$

so $\mathcal{R}_3 = O(h^2_{\max})$.

B.4 Exact oracle ranking under a margin

Let

$$\epsilon_{\max} = \max_k \left| \|\tau_k^E\| - s_k \right|,$$

and order the estimated scores as $s_{(1)} \geq \cdots \geq s_{(a)}$. Suppose $K < a$ and

$$s_{(K)} - s_{(K+1)} > 2\epsilon_{\max}.$$

For every selected index $i \in S$ and excluded index $j \notin S$,

$$\|\tau_i^E\| \geq s_i - \epsilon_{\max} \geq s_{(K)} - \epsilon_{\max} > s_{(K+1)} + \epsilon_{\max} \geq s_j + \epsilon_{\max} \geq \|\tau_j^E\|.$$

Thus every selected true defect norm is strictly larger than every excluded one, so S is the unique oracle top-K set. This completes the proof of Theorem 1.

C Endpoint reuse and exact-stage query complexity

C.1 Uniqueness of the endpoint-reusable RK2 method

A general explicit two-stage Runge–Kutta method has the form

$$k_1 = f(x,t), \qquad k_2 = f(x + chk_1, t + ch),$$
$$x^+ = x + h(b_1 k_1 + b_2 k_2). \tag{C.1}$$

Expanding the second stage around (x,t),

$$k_2 = k_1 + ch\big[J_x f(x,t) f(x,t) + \partial_t f(x,t)\big] + O(h^2). \tag{C.2}$$

Substitution into equation C.1 and comparison with the exact-flow expansion in equation B.1 give the second-order conditions

$$b_1 + b_2 = 1, \qquad b_2 c = \frac{1}{2}. \tag{C.3}$$

For the second-stage state–time point to coincide with the Euler endpoint for all x, h, f, we must have $c = 1$. Equation C.3 then gives

$$b_1 = b_2 = \frac{1}{2},$$

which is Heun's method.

C.2 Exact evaluation count

Let $p_1 < \cdots < p_K$ be the selected intervals. First suppose $p_1 \leq a - 2$, which includes every $K \geq 2$ and also a nonterminal singleton. No earlier state has changed, and therefore

$$y_{p_1+1} = \bar{x}_{p_1} + h_{p_1} \bar{v}_{p_1} = \bar{x}_{p_1+1}.$$

The required endpoint stage is exactly the recorded value $\bar{v}_{p_1+1}$, so the first selected interval uses no new query. Algorithm 1 evaluates one new endpoint stage for each of $p_2, \ldots, p_K$, giving exactly $K - 1$ new model evaluations. If $S = \{a - 1\}$, its endpoint velocity is absent from the Euler trace and one new evaluation is made.

C.3 Worst-case lower bound for exact endpoint stages

Let $\Xi_2(f), \ldots, \Xi_K(f) \in \mathbb{R}^d \times [0,1]$ denote the recursively generated endpoint-stage inputs after the first correction. Assume that, for a chosen base field, these $K - 1$ points are pairwise distinct and are absent from the reference query set. Suppose an adaptive black-box algorithm always returns all exact values

$$f(\Xi_2(f)), \ldots, f(\Xi_K(f))$$

for every admissible C^2 field while making at most $K - 2$ new oracle queries.

Choose a base field f_0 satisfying the stated distinctness conditions. Run the algorithm against f_0, and let $\mathcal{Q}_0$ be the finite union of the reference query set and all new query points used in this run. At least one required target is not in $\mathcal{Q}_0$. Let r be the first target index with

$$\Xi_r(f_0) \notin \mathcal{Q}_0. \tag{C.4}$$

Choose an open ball U centered at $\Xi_r(f_0)$ whose closure avoids $\mathcal{Q}_0$, a scalar bump function $\psi \in C_c^\infty(U)$ with $\psi(\Xi_r(f_0)) = 1$, and a nonzero vector u. For sufficiently small $\varepsilon \neq 0$, define

$$f_1(z) = f_0(z) + \varepsilon\psi(z)u. \tag{C.5}$$

Then f_0 and f_1 agree at every point in $\mathcal{Q}_0$. Their reference traces are identical, and an induction over the adaptive queries shows that the complete oracle transcript is identical. By the choice of r, all earlier required endpoint stages also agree; hence the recursively generated target remains unchanged:

$$\Xi_r(f_1) = \Xi_r(f_0). \tag{C.6}$$

At that point, however,

$$f_1(\Xi_r(f_1)) = f_0(\Xi_r(f_0)) + \varepsilon u \neq f_0(\Xi_r(f_0)). \tag{C.7}$$

The algorithm produces the same output under the two indistinguishable transcripts and therefore cannot be exact for both fields. This contradiction proves that at least $K - 1$ new queries are necessary in the worst case. Together with Appendix C.2, this proves the exact-stage query claim in Section 3.4.

D Proof of Theorem 2

D.1 Spatial information identified by a reference triplet

Fix three times $t_- < t_0 < t_+$ and let

$$\theta = \frac{t_0 - t_-}{t_+ - t_-}, \qquad t_0 = (1-\theta)t_- + \theta t_+.$$

Suppose the three observations satisfy an affine model

$$v_i = Ax_i + gt_i + \beta, \qquad i \in \{-, 0, +\}. \tag{D.1}$$

Define

$$q = x_0 - [(1-\theta)x_- + \theta x_+], \qquad r = v_0 - [(1-\theta)v_- + \theta v_+].$$

Taking the same time-chord residual in equation D.1 cancels $gt + \beta$, so every compatible spatial matrix satisfies

$$Aq = r. \tag{D.2}$$

Conversely, let A' satisfy $A'q = r$ and define $z_i = v_i - A'x_i$. Equation D.2 is equivalent to

$$z_0 = (1-\theta)z_- + \theta z_+. \tag{D.3}$$

There is a vector-valued affine function $g't + \beta'$ interpolating (t_-, z_-) and (t_+, z_+); equation D.3 ensures that it also equals z_0 at t_0. Hence

$$v_i = A'x_i + g't_i + \beta' \qquad i \in \{-, 0, +\}.$$

Therefore the compatible spatial matrices are exactly

$$\mathcal{A} = \{A' : A'q = r\}. \tag{D.4}$$

D.2 Minimum-norm extension

Assume $q \neq 0$ and define

$$B^\star = \frac{rq^\top}{q^\top q}.$$

Then $B^\star q = r$. For any other feasible B, let $H = B - B^\star$; then $Hq = 0$. The Frobenius inner product satisfies

$$\langle B^\star, H\rangle_F = \operatorname{tr}\left(\frac{qr^\top}{q^\top q}H\right) = \frac{r^\top Hq}{q^\top q} = 0.$$

Consequently,

$$\|B\|_F^2 = \|B^\star\|_F^2 + \|B - B^\star\|_F^2, \tag{D.5}$$

which proves uniqueness and minimum Frobenius norm (Dennis & Schnabel, 1979).

Every feasible B also satisfies

$$\|r\| = \|Bq\| \le \|B\|_2\|q\|,$$

so $\|B\|_2 \ge \|r\|/\|q\|$. Since $B^\star$ has rank at most one,

$$\|B^\star\|_2 = \frac{\|r\|}{\|q\|}, \tag{D.6}$$

and the lower bound is attained. Finally, if $q^\top z = 0$, then

$$B^\star z = r\frac{q^\top z}{q^\top q} = 0. \tag{D.7}$$

When $q = 0$, minimizing $\|Bq - r\|^2$ leaves all matrices tied, and $B^\star = 0$ is the unique one of minimum Frobenius norm. Under an exact affine model, $q = 0$ also implies $r = 0$.

D.3 Alignment and exact affine action

Fix any interior node $1 \le j \le a - 2$ and set

$$h_- = t_j - t_{j-1}, \qquad h_+ = t_{j+1} - t_j.$$

The reference Euler relations are

$$\bar{x}_{j-1} = \bar{x}_j - h_-\bar{v}_{j-1}, \qquad \bar{x}_{j+1} = \bar{x}_j + h_+\bar{v}_j. \tag{D.8}$$

Substituting equation D.8 into the time chord in the definition of q_j gives

$$q_j = -\frac{h_-h_+}{h_- + h_+}(\bar{v}_j - \bar{v}_{j-1}) = -\frac{2h_+}{h_- + h_+}\widetilde{d}_{j-1}. \tag{D.9}$$

If interval $j - 1$ is upgraded to Heun from its reference state, its displacement is

$$\delta_j = \frac{h_-}{2}(\bar{v}_j - \bar{v}_{j-1}) = \widetilde{d}_{j-1} = -\frac{h_- + h_+}{2h_+}q_j. \tag{D.10}$$

Thus $\delta_j \in \operatorname{span}(q_j)$.

For an affine field $f(x,t) = A_jx + g_jt + \beta_j$ on $\mathcal{N}_j$, the chord identity gives $r_j = A_jq_j$. If $q_j \ne 0$,

$$B_j^\star = \frac{A_jq_jq_j^\top}{q_j^\top q_j} = A_jP_j, \qquad P_j = \frac{q_jq_j^\top}{q_j^\top q_j}. \tag{D.11}$$

Equation D.10 gives $P_j\delta_j = \delta_j$, and hence

$$B_j^\star\delta_j = A_j\delta_j.$$

If the corrected state remains in $\mathcal{N}_j$, then

$$\bar{v}_j + B_j^\star\delta_j = A_j(\bar{x}_j + \delta_j) + g_jt_j + \beta_j = f(\bar{x}_j + \delta_j, t_j).$$

If $q_j = 0$, equation D.10 gives $\delta_j = 0$, so the same conclusion holds. This completes the proof of Theorem 2.

E PROOFS OF THE PROPAGATION GUARANTEES

E.1 EXACT AFFINE SEGMENTS

Assume a uniform step size h and an affine field

$$f(x,t) = Ax + gt + \beta$$

on a correction-free propagation segment and the reference triplets used there. We consider nodes $s,\dots,m$ for which the three-point secants are available, so $1 \le s \le m \le a-2$. The reference velocity recurrence is

$$\bar{v}_{j+1} = \bar{v}_j + h(A\bar{v}_j + g). \tag{E.1}$$

On a uniform grid, equation D.9 becomes

$$q_j = -\frac{h}{2}(\bar{v}_j - \bar{v}_{j-1}) = -\frac{h^2}{2}(A\bar{v}_{j-1} + g). \tag{E.2}$$

Equation E.1 implies

$$A\bar{v}_j + g = (I + hA)(A\bar{v}_{j-1} + g),$$

and therefore

$$q_{j+1} = (I + hA)q_j. \tag{E.3}$$

Suppose $\delta_s = \alpha q_s$. Equation D.11 gives $B_s^\star \delta_s = A\delta_s$, so

$$v_s = \bar{v}_s + A\delta_s = f(x_s, t_s). \tag{E.4}$$

With no new correction injected on interval s,

$$\delta_{s+1} = (I + hA)\delta_s = \alpha(I + hA)q_s = \alpha q_{s+1}. \tag{E.5}$$

Induction gives $\delta_j = \alpha q_j$ and $v_j = f(x_j, t_j)$ throughout the segment; if a chord becomes zero, its aligned displacement also becomes zero and the induction continues. Every update whose left endpoint admits the three-point secant therefore matches full velocity re-evaluation. This proves the affine-segment statement in Section 4.1.

E.2 PROOF OF PROPOSITION 1

Write $h_- = h_{j-1}$ and $h_+ = h_j$. Abbreviate

$$\begin{aligned} \Delta x^- &= \bar{x}_{j-1} - \bar{x}_j, \qquad \Delta t^- = t_{j-1} - t_j, \\ \Delta x^+ &= \bar{x}_{j+1} - \bar{x}_j, \qquad \Delta t^+ = t_{j+1} - t_j, \\ \Delta z_j^\pm &= (\Delta x^\pm, \Delta t^\pm). \end{aligned}$$

Taylor expansion of the neighboring velocities around $(\bar{x}_j, t_j)$ gives

$$\bar{v}_{j-1} = \bar{v}_j + J_j \Delta x^- + \partial_t f(\bar{x}_j, t_j)\Delta t^- + R^-, \tag{E.6}$$

$$\bar{v}_{j+1} = \bar{v}_j + J_j \Delta x^+ + \partial_t f(\bar{x}_j, t_j)\Delta t^+ + R^+, \tag{E.7}$$

where

$$\|R^-\| \le \frac{M}{2}\|\Delta z_j^-\|^2, \qquad \|R^+\| \le \frac{M}{2}\|\Delta z_j^+\|^2. \tag{E.8}$$

The choice of θ_j gives

$$(1-\theta_j)\Delta t^- + \theta_j \Delta t^+ = 0,$$

while the definition of q_j gives

$$(1-\theta_j)\Delta x^- + \theta_j \Delta x^+ = -q_j.$$

Substituting equation E.6–equation E.7 into the velocity-chord residual yields

$$r_j = J_j q_j - \big[(1-\theta_j)R^- + \theta_j R^+\big]. \tag{E.9}$$

Hence

$$\varepsilon_j = r_j - J_j q_j = -\big[(1-\theta_j)R^- + \theta_j R^+\big],$$

and equation E.8 proves the stated chord-residual bound.

For $q_j \neq 0$,

$$B_j^\star = \frac{(J_j q_j + \varepsilon_j) q_j^\top}{q_j^\top q_j} = J_j P_j + \varepsilon_j \frac{q_j^\top}{q_j^\top q_j}. \tag{E.10}$$

At fixed time t_j, spatial Taylor expansion gives

$$f(\bar{x}_j + \delta_j, t_j) = \bar{v}_j + J_j \delta_j + R_j(\delta_j), \qquad \|R_j(\delta_j)\| \leq \frac{M}{2} \|\delta_j\|^2. \tag{E.11}$$

Subtracting $\bar{v}_j + B_j^\star \delta_j$ and using equation E.10 proves

$$f(\bar{x}_j + \delta_j, t_j) - (\bar{v}_j + B_j^\star \delta_j) = J_j (I - P_j)\delta_j - \varepsilon_j \frac{q_j^\top \delta_j}{q_j^\top q_j} + R_j(\delta_j). \tag{E.12}$$

Taking norms and using

$$\|P_j \delta_j\| = \frac{|q_j^\top \delta_j|}{\|q_j\|}$$

gives the bound in Proposition 1. If $q_j = 0$, the method sets $B_j^\star = 0$, and equation E.11 directly gives

$$\eta_j \leq \|J_j\| \, \|\delta_j\| + \frac{M}{2} \|\delta_j\|^2. \tag{E.13}$$

At an interior node following a Heun correction from its reference state, Theorem 2 gives

$$\delta_j = \alpha_j q_j, \qquad \alpha_j = -\frac{h_- + h_+}{2h_+}. \tag{E.14}$$

Let M_f bound the reference velocities. The Euler relations give $\|\Delta z_j^-\|^2 \leq (1 + M_f^2) h_-^2$ and $\|\Delta z_j^+\|^2 \leq (1 + M_f^2) h_+^2$. Hence the weighted Taylor bound is

$$\begin{aligned} \|\varepsilon_j\| &\leq \frac{M(1 + M_f^2)}{2} \left(\frac{h_+ h_-^2 + h_- h_+^2}{h_- + h_+} \right) \\ &= C_\varepsilon h_- h_+, \qquad C_\varepsilon = \frac{M(1 + M_f^2)}{2}. \end{aligned} \tag{E.15}$$

Since $|\alpha_j| = (h_- + h_+)/(2h_+)$, $|\alpha_j| \|\varepsilon_j\| \leq (C_\varepsilon/2) h_- (h_- + h_+)$. The reference stage expansion gives $\|\delta_j\| = O(h_-^2)$. Since $(I - P_j)\delta_j = 0$, the extension decomposition yields

$$\begin{aligned} \eta_j &\leq |\alpha_j| \|\varepsilon_j\| + \frac{M}{2} \|\delta_j\|^2 \\ &\leq \frac{C_\varepsilon}{2} h_- (h_- + h_+) + C h_-^4 = O\big(h_- (h_- + h_+)\big) = O(h_{\max}^2). \end{aligned} \tag{E.16}$$

Here $h_- \leq 1$, so the fourth-order term is absorbed.

For $q_j \neq 0$, equation E.14 also gives

$$B_j^\star \delta_j = \alpha_j r_j = -\frac{h_- + h_+}{2h_+} r_j. \tag{E.17}$$

If $q_j = 0$, equation E.14 gives $\delta_j = 0$, and the same error bound follows from the zero-chord fallback $B_j^\star = 0$.

If such a replacement ends at the boundary velocity node $a - 1$, then $v_{a-1} = \bar{v}_{a-1}$ and $\delta_{a-1} = \widetilde{d}_{a-2} = O(h_{\max}^2)$. Lipschitz continuity therefore gives

$$\eta_{a-1} = \|f(\bar{x}_{a-1} + \delta_{a-1}, t_{a-1}) - f(\bar{x}_{a-1}, t_{a-1})\| = O(h_{\max}^2). \tag{E.18}$$

This proves Proposition 1.

E.3 A POSTERIORI DISPLACEMENT STABILITY

From the displacement recursion,

$$\begin{aligned}\|\delta_{j+1}\| &\le \|\delta_j\| + h_j\|B_j^\star\delta_j\| + \mathbf{1}_{\{j\in S\}}\|u_{j+1}\| \\ &= (1+h_j\gamma_j)\|\delta_j\| + \mathbf{1}_{\{j\in S\}}\|u_{j+1}\|.\end{aligned} \tag{E.19}$$

Iterating equation E.19 from s to m produces products of factors $1+h_j\gamma_j$. Since $1+z\le e^z$ for $z\ge 0$, each product is at most

$$\exp\left(\sum_{j=s}^{m-1} h_j\gamma_j\right).$$

Factoring out this common upper bound proves equation 13. Since $\gamma_j \le \|B_j^\star\|_2$ and $\sum_j h_j \le 1$, the factor is at most $\exp(\max_j \|B_j^\star\|_2)$.

F END-TO-END ANALYSIS

F.1 REFERENCE-TO-CURRENT DEFECT TRANSFER

For an arbitrary current state x at time t_k, define

$$\tau_k^E(x) = \Phi_{t_k,h_k}(x) - [x + h_k f(x,t_k)].$$

Uniformly on the compact neighborhood, the expansion in equation B.2 gives

$$\tau_k^E(x) = \frac{h_k^2}{2}\mathcal{M}(x,t_k) + R_k(x), \qquad \|R_k(x)\| \le C_R h_k^3. \tag{F.1}$$

For $k \le a-2$, equation B.4 gives

$$\widetilde{d}_k = \frac{h_k^2}{2}\mathcal{M}(\bar{x}_k,t_k) + \widetilde{R}_k, \qquad \|\widetilde{R}_k\| \le C_D h_k^3. \tag{F.2}$$

With $x_k = \bar{x}_k + \delta_k$ and the Lipschitz continuity of $\mathcal{M}$,

$$\|\tau_k^E(x_k) - \widetilde{d}_k\| \le C_1 h_k^3 + C_2 h_k^2\|\delta_k\| \qquad (k \le a-2). \tag{F.3}$$

For $k = a-1$, insert and subtract the reference defect and use equation B.11:

$$\|\tau_{a-1}^E(x_{a-1}) - \widetilde{d}_{a-1}\| \le C_1 h_{a-1}^3 + C_2 h_{a-1}^2\|\delta_{a-1}\| + C_\partial h_{a-1}^2 h_{a-2}. \tag{F.4}$$

Combining the interior and terminal cases gives

$$\|\tau_k^E(x_k) - \widetilde{d}_k\| \le C_1 h_k^3 + C_2 h_k^2\|\delta_k\| + \mathbf{1}_{\{k=a-1\}} C_\partial h_{a-1}^2 h_{a-2}. \tag{F.5}$$

If an unselected Euler update uses v_k instead of $f(x_k,t_k)$, then

$$\begin{aligned}&\|\Phi_{t_k,h_k}(x_k) - (x_k + h_k v_k)\| \\ &\quad\le \|\widetilde{d}_k\| + C_1 h_k^3 + C_2 h_k^2\|\delta_k\| + h_k\eta_k + \mathbf{1}_{\{k=a-1\}} C_\partial h_{a-1}^2 h_{a-2}.\end{aligned} \tag{F.6}$$

F.2 INEXACT-STAGE HEUN AND ENDPOINT ERROR DECOMPOSITION

Let $v = f(x,t) + e$, with $\|e\| = \eta$, and define

$$y = x + hv, \qquad \widetilde{H}_h(x) = x + \frac{h}{2}[v + f(y, t+h)].$$

Let $v^\star = f(x,t)$, $y^\star = x + hv^\star$, and

$$H_h(x) = x + \frac{h}{2}[v^\star + f(y^\star, t+h)]$$

be exact-stage Heun. Since $\|y - y^\star\| \le h\eta$ and f is L-Lipschitz,

$$\|\widetilde{H}_h(x) - H_h(x)\| \le \frac{h}{2}(1 + Lh)\eta. \tag{F.7}$$

Exact-stage Heun has local error $\|\Phi_{t,h}(x) - H_h(x)\| \le C_H h^3$, by the expansions in Appendix B.1 applied at x.

Let $\ell_k := \|\Phi_{t_k,h_k}(x_k) - x_{k+1}\|$. The preceding estimates give

$$\ell_k \le \begin{cases} \|\widetilde{d}_k\| + C_1 h_k^3 + C_2 h_k^2 \|\delta_k\| + h_k \eta_k + \mathbf{1}_{\{k=a-1\}} C_\partial h_{a-1}^2 h_{a-2}, & k \notin S, \\ C_H h_k^3 + \dfrac{h_k}{2}(1 + Lh_k)\eta_k, & k \in S. \end{cases} \tag{F.8}$$

Let $x_k^\star = x^\star(t_k)$ be the exact trajectory from the same initial state and set $e_k = x_k^\star - x_k$. Let $\Psi_k^{\mathrm{Fix}}(x_k) = x_{k+1}$ denote the actual update, with the reference trace and selected set fixed. Inserting the exact flow from the current numerical state,

$$\begin{aligned} e_{k+1} &= \Phi_{t_k,h_k}(x_k^\star) - \Psi_k^{\mathrm{Fix}}(x_k) \\ &= [\Phi_{t_k,h_k}(x_k^\star) - \Phi_{t_k,h_k}(x_k)] + [\Phi_{t_k,h_k}(x_k) - \Psi_k^{\mathrm{Fix}}(x_k)]. \end{aligned} \tag{F.9}$$

Continuous dependence of the exact flow gives

$$\|\Phi_{t_k,h_k}(x) - \Phi_{t_k,h_k}(z)\| \le e^{Lh_k}\|x - z\|. \tag{F.10}$$

Using equation F.8 and equation F.10,

$$\|e_{k+1}\| \le e^{Lh_k}\|e_k\| + \ell_k, \tag{F.11}$$

where ℓ_k is the corresponding local bound. Iterating from $e_0 = 0$ and using $\sum_k h_k = 1$,

$$\|e_a\| \le e^L \sum_{k=0}^{a-1} \ell_k. \tag{F.12}$$

The unresolved-defect contributions sum to

$$\sum_{k \notin S} \|\widetilde{d}_k\|.$$

All cubic and final-boundary remainders are bounded by a constant multiple of $\mathcal{R}_3$. The reference-to-current terms sum to

$$C_\delta \sum_{k \notin S} h_k^2 \|\delta_k\|,$$

and the velocity-extension terms are bounded by

$$C_\eta \sum_{k=0}^{a-1} h_k \eta_k.$$

Absorbing the flow-stability factor and the grid-independent coefficients into a common constant C proves the error decomposition in Theorem 3. Since S contains the K largest $\|\widetilde{d}_k\|$, it exactly minimizes the first term over all selected sets of size K.

F.3 Proof of Corollary 1

Assume $\|B_j^\star\|_2 \le M_B$ uniformly. We first derive bounds in terms of $Q_S := \sum_{k \in S} h_k^2$, without fixing K. Set $B_0^\star = B_{a-1}^\star = 0$. At each velocity node,

$$v_j = \bar{v}_j + B_j^\star \delta_j.$$

Let M_f bound $\|f\|$, and let L_x, L_t be state and time Lipschitz constants. Then

$$\|v_j\| \le M_f + M_B\|\delta_j\|, \tag{F.13}$$

and

$$\eta_j = \|f(\bar{x}_j + \delta_j, t_j) - f(\bar{x}_j, t_j) - B_j^\star \delta_j\| \le (L_x + M_B)\|\delta_j\|. \tag{F.14}$$

At a selected interval,

$$u_{j+1} = \frac{h_j}{2}[f(x_j + h_j v_j, t_j + h_j) - v_j].$$

Inserting $f(x_j, t_j)$ and applying equation F.13–equation F.14 gives constants C_u, C_u', independent of the grid, such that

$$\|u_{j+1}\| \le C_u h_j^2 + C_u' h_j \|\delta_j\|. \tag{F.15}$$

Set $D_j = \|\delta_j\|$ and $D = \max_j D_j$. By the displacement recursion and equation F.15,

$$D_{j+1} \le (1 + C_0 h_j) D_j + C_u \mathbf{1}_{\{j \in S\}} h_j^2, \qquad C_0 = M_B + C_u'. \tag{F.16}$$

Iterating from $D_0 = 0$ and using $\sum_j h_j = 1$ gives

$$D \le C_u e^{C_0} \sum_{j \in S} h_j^2 = C_u e^{C_0} Q_S. \tag{F.17}$$

Equation F.14 then gives $\max_j \eta_j \le C Q_S$. The constants are independent of K, the selected set, and the grid, subject to the stated uniform bounds. In particular, for fixed K, $Q_S \le K h_{\max}^2$ implies

$$D = O(h_{\max}^2), \qquad \max_j \eta_j = O(h_{\max}^2). \tag{F.18}$$

We now assume K fixed to derive the remaining claims of Corollary 1.

For a nonterminal selected interval, compare the actual correction

$$u_{k+1} = \frac{h_k}{2}[f(x_k + h_k v_k, t_{k+1}) - v_k]$$

with the reference correction

$$\widetilde{d}_k = \frac{h_k}{2}[f(\bar{x}_k + h_k \bar{v}_k, t_{k+1}) - \bar{v}_k].$$

Using $v_k - \bar{v}_k = B_k^\star \delta_k$, Lipschitz continuity gives

$$\|u_{k+1} - \widetilde{d}_k\| \le \frac{h_k}{2}\big[L_x(1 + h_k M_B) + M_B\big]\|\delta_k\| = O(h_{\max}^3). \tag{F.19}$$

For the final interval, introduce the exact reference Heun correction

$$d_{a-1}^{\mathrm{end}} = \frac{h_{a-1}}{2}[f(\bar{x}_{a-1} + h_{a-1}\bar{v}_{a-1}, t_a) - \bar{v}_{a-1}].$$

The same calculation gives $\|u_a - d_{a-1}^{\mathrm{end}}\| = O(h_{\max}^3)$, while Appendix B gives

$$\|d_{a-1}^{\mathrm{end}} - \widetilde{d}_{a-1}\| = O(h_{\max}^3).$$

Thus equation F.19 holds for every selected interval.

Substituting equation F.18 into the transfer bound in Appendix F.1 gives

$$\max_k \|\tau_k^E(x_k) - \widetilde{d}_k\| = O(h_{\max}^3). \tag{F.20}$$

For this realized current trajectory, set $d_k^{\mathrm{cur}} = \|\tau_k^E(x_k)\|$. The reverse triangle inequality gives $|d_k^{\mathrm{cur}} - s_k| \le C_{\mathrm{cur}} h_{\max}^3$. If $K < a$ and $s_{(K)} - s_{(K+1)} > 2C_{\mathrm{cur}} h_{\max}^3$, then for $i \in S$ and $j \notin S$,

$$d_i^{\mathrm{cur}} \ge s_i - C_{\mathrm{cur}} h_{\max}^3 > s_j + C_{\mathrm{cur}} h_{\max}^3 \ge d_j^{\mathrm{cur}}.$$

Thus S is also the unique top-K set of current-state Euler defect norms, as asserted in Corollary 1.

Even without this gap, equation B.13 applies to the fixed list $d_k^{\mathrm{cur}} = \|\tau_k^E(x_k)\|$ and the original selected scores. Since $|d_k^{\mathrm{cur}} - s_k| \le C_{\mathrm{cur}} h_{\max}^3$,

$$\sum_{k \notin S} d_k^{\mathrm{cur}} \le \min_{|A|=K} \sum_{k \notin A} d_k^{\mathrm{cur}} + 2K C_{\mathrm{cur}} h_{\max}^3. \tag{F.21}$$

For the comparison with the ideal selective Euler–Heun rerun, we use the general Q_S bounds above, without fixing K. Define

$$\Psi_k(x) = \begin{cases} x + h_k f(x, t_k), & k \notin S, \\ x + \dfrac{h_k}{2}\big[f(x, t_k) + f(x + h_k f(x, t_k), t_{k+1})\big], & k \in S, \end{cases} \tag{F.22}$$

and $z_0 = \bar{x}_0$, $z_{k+1} = \Psi_k(z_k)$. Then $z_j = x_j^{\mathrm{ideal}(S)}$ for every $0 \le j \le a$. For the same input state x_k, the actual FixUp step differs from $\Psi_k(x_k)$ by at most $Ch_k\eta_k$, for both Euler and Heun intervals. The Euler map has Lipschitz factor $1+L_x h_k$; the Heun map has factor $1+L_x h_k+\frac{1}{2}L_x^2 h_k^2$. Since $h_k \le 1$, both are bounded by $1 + C_\Psi h_k$, with a grid-independent C_Ψ. Therefore, with $\rho_k = \|x_k - z_k\|$,

$$\rho_{k+1} \le (1 + C_\Psi h_k)\rho_k + Ch_k\eta_k. \tag{F.23}$$

Iterating equation F.23 from $\rho_0 = 0$ to any $0 \le m \le a$ gives $\rho_m \le C\sum_{k=0}^{m-1} h_k\eta_k$, with the same constant for every prefix since $\sum_{k=0}^{m-1} h_k \le 1$. Consequently,

$$\max_{0\le j\le a} \|x_j - x_j^{\mathrm{ideal}(S)}\| \le C\sum_{k=0}^{a-1} h_k\eta_k \le CQ_S. \tag{F.24}$$

Together with equation F.17 and the extension bound, this proves the budget-aware bounds used in the main text. For fixed K, it gives

$$\max_{0\le j\le a} \|x_j - x_j^{\mathrm{ideal}(S)}\| = O(h_{\max}^2). \tag{F.25}$$

Equations F.18–F.25 prove Corollary 1 and the uniform ideal-rerun comparison in Section 4.2.

F.4 Affine rerun agreement

Proposition 3 (Affine rerun agreement)**.** Suppose $f(x,t) = Ax + gt + \beta$ on the common neighborhood, with constant A, g, β, and use a uniform grid with $h = 1/a$. For any selected set S, $K = |S|$, FixUp-FM and the ideal selective Euler–Heun rerun on the same grid and selected set satisfy

$$\max_{0\le j\le a} \|x_j - x_j^{\mathrm{ideal}(S)}\| \le C\big(Kh^3 + K^2h^4\big), \tag{F.26}$$

where C is independent of the grid, S, and K. For fixed K, the agreement is third order.

Proof. Put $M = I + hA$ and $\chi_j = \mathbf{1}_{\{j\in S\}}$. Extend the chord sequence algebraically by

$$q_j = -\frac{h^2}{2}(A\bar{v}_{j-1} + g), \qquad 1 \le j \le a.$$

This agrees with equation E.2 at interior nodes. The reference velocity recurrence gives $q_{j+1} = Mq_j$ and $\max_j \|q_j\| \le Ch^2$. Define the exact reference Heun increment $d_j^H := \frac{h^2}{2}(A\bar{v}_j + g) = -q_{j+1}$ for $0 \le j < a$. Let $n_j = |\{k \in S : k < j\}|$ and set

$$w_0 = 0, \qquad w_{j+1} = Mw_j + \chi_j d_j^H.$$

Induction yields $w_j = -n_j q_j$ for $1 \le j \le a$, so $\max_j \|w_j\| \le CKh^2$. At interior nodes, $B_j^\star q_j = Aq_j$ and $\|B_j^\star\|_2 \le \|A\|_2$, including the zero-chord case. Hence $B_j^\star w_j = Aw_j$.

For a displacement z and a velocity map B, the affine Heun correction is

$$U_j(z) = d_j^H + \frac{h}{2}(A - B)z + \frac{h^2}{2}ABz.$$

Thus, writing $\delta_j^I = x_j^{\mathrm{ideal}(S)} - \bar{x}_j$,

$$\begin{aligned} \delta_{j+1} &= T_j(B_j^\star)\delta_j + \chi_j d_j^H, \\ \delta_{j+1}^I &= T_j(A)\delta_j^I + \chi_j d_j^H, \\ T_j(B) &= I + hB + \chi_j\left[\frac{h}{2}(A - B) + \frac{h^2}{2}AB\right]. \end{aligned} \tag{F.27}$$

Both update matrices have norm at most $1 + Ch$, so their products over any grid segment are uniformly bounded. For $0 \le j \le a-2$,

$$T_j(B_j^\star)w_j - Mw_j = T_j(A)w_j - Mw_j = \frac{\chi_j h^2}{2} A^2 w_j,$$

where the $j = 0$ identity follows from $w_0 = 0$. Subtracting the recurrence for w_j from equation F.27 and summing these at most K forcing terms gives

$$\max_{0\le j\le a-1} \big(\|\delta_j - w_j\| + \|\delta_j^I - w_j\|\big) \le CK^2h^4. \tag{F.28}$$

At the final velocity node, $B_{a-1}^\star = 0$ and

$$[T_{a-1}(0) - M]w_{a-1} = -\left(1 - \frac{\chi_{a-1}}{2}\right) hAw_{a-1},$$

whose norm is at most CKh^3. The corresponding ideal forcing is $\chi_{a-1}h^2A^2w_{a-1}/2$. Combining this last update with equation F.28 proves equation F.26. □

Numerical examples. We use float64 with $a = 16, 32, 64, 128, 256, 512$ uniform steps. The scalar example is $\dot{x} = -x$, $x(0) = 1$, with $K = 1$; defect-based selection chooses $S = \{0\}$. The two-dimensional example is $\dot{x} = Ax + gt + \beta$, with

$$A = \begin{pmatrix} 0.7 & 1.2 \\ -0.3 & -0.4 \end{pmatrix}, \quad g = \begin{pmatrix} 0.2 \\ -0.1 \end{pmatrix}, \quad \beta = \begin{pmatrix} 0.1 \\ 0.3 \end{pmatrix}, \quad x(0) = \begin{pmatrix} 1 \\ 0.5 \end{pmatrix},$$

and $S = \{a/8, 3a/8, 5a/8\}$. In both examples, minimum-norm propagation, unchanged reference velocities, and the ideal rerun use the same grid and selected set. Minimum-norm propagation uses the original triplet construction and boundary maps $B_0^\star = B_{a-1}^\star = 0$. Table 2 reports the endpoint differences $E_a = \|x_a - x_a^{\mathrm{ideal}(S)}\|_2$.

Table 2: Endpoint differences from the ideal selective Euler–Heun rerun on the affine examples. Unchanged uses the reference velocities. The last row gives $p = \log_2(E_{256}/E_{512})$.

	Scalar, $K = 1$		Two-dimensional, $K = 3$	
a	Minimum-norm	Unchanged	Minimum-norm	Unchanged
16	4.9455×10^{-5}	1.2113×10^{-3}	4.9142×10^{-5}	4.5268×10^{-4}
32	5.8867×10^{-6}	3.0579×10^{-4}	5.9973×10^{-6}	1.2221×10^{-4}
64	7.1843×10^{-7}	7.6809×10^{-5}	7.4037×10^{-7}	3.1658×10^{-5}
128	8.8747×10^{-8}	1.9247×10^{-5}	9.1958×10^{-8}	8.0511×10^{-6}
256	1.1028×10^{-8}	4.8172×10^{-6}	1.1458×10^{-8}	2.0298×10^{-6}
512	1.3745×10^{-9}	1.2050×10^{-6}	1.4299×10^{-9}	5.0956×10^{-7}
p	3.004	1.999	3.002	1.994

F.5 Convergence when all intervals are selected

Suppose every interval is selected and $\max_k \eta_k = O(h_{\max}^2)$. The unresolved-defect and reference-to-current terms in Theorem 3 vanish. Since $\sum_k h_k = 1$,

$$\sum_k h_k \eta_k = O(h_{\max}^2), \qquad \mathcal{R}_3 = O(h_{\max}^2).$$

Theorem 3 therefore gives

$$\|x^\star(1) - x_a\| = O(h_{\max}^2). \tag{F.29}$$

G STABILITY AND CONSISTENCY OF ANYTIME CONTINUATION

We prove Proposition 2 for the continuation rule in Section 3.4. Throughout, the reference trace and operators $B_j^\star$ remain fixed, each interval is corrected at most once, and a previously applied correction is retained when subsequent requests propagate through that interval. We assume the uniform regularity and propagation conditions of Corollary 1. For any interval set A, write

$$Q_A := \sum_{k\in A} h_k^2, \qquad H_A := \sum_{k\in A} h_k.$$

A common propagation representation. At the end of any completed request, let c_k be the correction first applied on interval k, with $c_k = 0$ for intervals not yet corrected. Subtracting the reference Euler update from the continuation update gives

$$\delta_{k+1} = (I + h_k B_k^\star)\delta_k + c_k, \qquad \delta_0 = 0. \tag{G.1}$$

The same relation holds on every updated prefix during a request. Define $F_{m,m} = I$ and, for $\ell < m$,

$$F_{m,\ell} := (I + h_{m-1}B_{m-1}^\star)\cdots(I + h_\ell B_\ell^\star).$$

Then

$$\delta_m = \sum_{k=0}^{m-1} F_{m,k+1}c_k, \qquad \|F_{m,\ell}\|_2 \le e^{M_B}, \tag{G.2}$$

since $\|B_j^\star\|_2 \le M_B$ and the integration horizon is one.

Budget-sensitive displacement control. For a displacement z at interval k, define the Heun correction

$$U_k(z) := \frac{h_k}{2}\left[f(\bar{x}_k + z + h_k(\bar{v}_k + B_k^\star z), t_{k+1}) - (\bar{v}_k + B_k^\star z)\right].$$

The same Lipschitz estimates used in Appendix F.3 give grid-independent constants such that

$$\|U_k(z)\| \le C_0 h_k^2 + C_1 h_k\|z\|, \qquad \|U_k(z) - U_k(z')\| \le C_2 h_k\|z - z'\|. \tag{G.3}$$

Let S be the final selected set of a continuation run, and list its intervals in the order their corrections are first formed as $k_1, \dots, k_N$. Set $G_r := \sum_{i=1}^r \|c_{k_i}\|$, with $G_0 = 0$. Before correction r is formed, equation G.2 bounds its input displacement by $e^{M_B}G_{r-1}$. Hence

$$G_r \le (1 + C_1 e^{M_B} h_{k_r})G_{r-1} + C_0 h_{k_r}^2.$$

Iterating and using $\sum_{k\in S} h_k \le 1$ yields

$$G_N \le CQ_S, \qquad \max_{\text{states encountered}} \|\delta\| \le CQ_S, \tag{G.4}$$

with C independent of the number, grouping, and ordering of the selected intervals. The velocity-extension estimate in Appendix F.3 therefore also gives $\eta \le CQ_S$ at every state encountered by the continuation procedure.

Comparison with a one-shot run. Let T be the selected set of a one-shot run from the same reference trace. Applying equation G.4 to the continuation and one-shot runs and using the triangle inequality at every node gives

$$\max_{0\le j\le a} \|x_j^{\text{any}} - x_j^{\text{one}}\| \le C(Q_S + Q_T). \tag{G.5}$$

If the two runs select the same set $S = T$, let z_k^{any} and z_k^{one} be their respective displacements when the correction on interval k is first formed. Equation G.4 gives $\|z_k^{\text{any}} - z_k^{\text{one}}\| \le CQ_S$. By equation G.3,

$$\|c_k^{\text{any}} - c_k^{\text{one}}\| \le Ch_k Q_S.$$

Because both runs use the same propagation operators, for every $0 \le m \le a$,

$$x_m^{\text{any}} - x_m^{\text{one}} = \sum_{\substack{k\in S\\ k<m}} F_{m,k+1}(c_k^{\text{any}} - c_k^{\text{one}}).$$

The norm is at most $CQ_S \sum_{k\in S,\, k<m} h_k \le CQ_S H_S$, so

$$\max_{0\le j\le a} \|x_j^{\mathrm{any}} - x_j^{\mathrm{one}}\| \le CQ_S H_S \qquad (S = T). \tag{G.6}$$

For a fixed total of K corrections, $Q_S, Q_T \le Kh_{\max}^2$ and $H_S \le Kh_{\max}$. Equations G.5–G.6 thus give the $O(h_{\max}^2)$ different-set and $O(h_{\max}^3)$ same-set rates in Proposition 2, uniformly over all trajectory nodes, request groupings, and correction orderings.

H Accuracy of refreshed anytime scores

At the start of a new request, the interval scores are recomputed by replacing the reference velocities with the current propagated velocities. Define

$$\widetilde{d}_k^{\mathrm{cur}} := \frac{h_k}{2}(v_{k+1} - v_k), \qquad 0 \le k \le a-2,$$

and

$$\widetilde{d}_{a-1}^{\mathrm{cur}} := \frac{h_{a-1}^2}{2h_{a-2}}(v_{a-1} - v_{a-2}), \qquad s_k^{\mathrm{cur}} := \|\widetilde{d}_k^{\mathrm{cur}}\|_2.$$

Under the same uniform regularity and propagation assumptions, assume the total number of corrections across requests is fixed as $h_{\max} \to 0$. Then equation G.4 gives $\max_j \|\delta_j\| = O(h_{\max}^2)$ at every completed-request scoring snapshot.

For a nonterminal interval,

$$\widetilde{d}_k^{\mathrm{cur}} - \widetilde{d}_k = \frac{h_k}{2}\left(B_{k+1}^\star \delta_{k+1} - B_k^\star \delta_k\right) = O(h_{\max}^3). \tag{H.1}$$

Combining this with equation F.5 and the fixed-budget displacement bound gives

$$\|\widetilde{d}_k^{\mathrm{cur}} - \tau_k^E(x_k)\| = O(h_{\max}^3), \qquad 0 \le k \le a-2. \tag{H.2}$$

The reverse triangle inequality therefore gives

$$\left|s_k^{\mathrm{cur}} - \|\tau_k^E(x_k)\|\right| = O(h_{\max}^3) \qquad (k \le a-2).$$

For the terminal interval, the boundary fallback $B_{a-1}^\star = 0$ yields

$$\widetilde{d}_{a-1}^{\mathrm{cur}} - \widetilde{d}_{a-1} = -\frac{h_{a-1}^2}{2h_{a-2}} B_{a-2}^\star \delta_{a-2}. \tag{H.3}$$

Thus the same $O(h_{\max}^3)$ rate holds at the terminal interval whenever h_{a-1}/h_{a-2} remains uniformly bounded. Under this condition, the refreshed score approximates the current terminal Euler defect norm to third order as well.

Consequently, the allocation and score-margin arguments of Appendix B apply to the remaining candidate intervals at each scoring snapshot. If the terminal interval is no longer a candidate, the additional step-ratio condition is unnecessary.

Near-oracle allocation on the final trajectory. Assume h_{a-1}/h_{a-2} is uniformly bounded, and let S be the final selected set, with $K = |S|$ fixed as $h_{\max} \to 0$. Define $d_k^{\mathrm{fin}} := \|\tau_k^E(x_k^{\mathrm{any}})\|$ on the final corrected trajectory. At scoring snapshot r, write $s_k^{(r)}$ for the refreshed score of an available interval. By equation G.4, the displacement is bounded by CQ_S both at every scoring snapshot and on the final trajectory. Combining equation H.1, equation H.3, and equation F.5 at the final states gives the common bound

$$\left|s_k^{(r)} - d_k^{\mathrm{fin}}\right| \le E_k := C\left[h_k^3 + \mathbf{1}_{\{k=a-1\}} h_{a-1}^2 h_{a-2} + h_k Q_S\right]. \tag{H.4}$$

Here $h_k^2 Q_S \le h_k Q_S$, and the terminal score perturbation is bounded by $Ch_{a-1}Q_S$ using the step-ratio assumption. The constant is independent of the snapshot, selected set, grid, and request grouping.

Let $A_\star$ be an oracle top-K set for the fixed list $\{d_k^{\mathrm{fin}}\}$. If $i \in S$ was selected at snapshot r and $j \notin S$, then j was still available at that snapshot. Selection of the highest refreshed scores therefore gives $s_i^{(r)} \geq s_j^{(r)}$, and hence $d_j^{\mathrm{fin}} - d_i^{\mathrm{fin}} \leq E_i + E_j$. Pairing the equally many indices in $A_\star \setminus S$ and $S \setminus A_\star$, and writing $W = A_\star \triangle S$, yields

$$\begin{aligned} \sum_{k \notin S} d_k^{\mathrm{fin}} - \min_{|A|=K} \sum_{k \notin A} d_k^{\mathrm{fin}} &\leq \sum_{k \in W} E_k \\ &\leq C\left(\mathcal{R}_3 + Q_S H_W\right). \end{aligned} \tag{H.5}$$

Since $|W| \leq 2K$ and $Q_S \leq K h_{\max}^2$, the same error sum is also at most $CK(1+K)h_{\max}^3$. Moreover, Hölder's inequality gives

$$Q_S H_W \leq K^{1/3} |W|^{2/3} \left(\sum_{k \in S} h_k^3\right)^{2/3} \left(\sum_{k \in W} h_k^3\right)^{1/3} \leq 2^{2/3} K \mathcal{R}_3.$$

Combining the two bounds proves

$$\sum_{k \notin S} \|\tau_k^E(x_k^{\mathrm{any}})\| \leq \min_{|A|=K} \sum_{k \notin A} \|\tau_k^E(x_k^{\mathrm{any}})\| + C_K \min\{\mathcal{R}_3, K h_{\max}^3\}. \tag{H.6}$$

The constant C_K may depend on the fixed total correction count, but not on the grid, selected set, request grouping, or correction ordering. In particular, the excess unresolved defect is $O(K h_{\max}^3)$ for fixed K, proving the anytime selection claim in Section 3.4.

I EXPERIMENTAL SETUP AND EVALUATION

Models, prompts, and pairing. We use SD3 Medium, SD3.5 Large, and FLUX.1-dev on the first 5,000 `user_input` entries of FluxPrompting's training split (revision `d137f642ba5e6252f4eefd6dd81500de5aedcebd`), in source order and without prompt rewriting. Sample indices are zero-based. The initial noise is generated on CPU with seed equal to the prompt index and reused across methods within each model and resolution. Anytime evaluation uses indices 0–2999. Refinement calibration uses the disjoint indices 5000–5063; adaptive-budget calibration uses indices 5000–5127, with its coarse-tolerance extension using the first 64 of these prompts.

Operating-point selection uses 1,024 validation prompts shared across models and disjoint from the evaluation cohort. They are taken in source order from rows 10000–11025 after removing empty entries and duplicate normalized texts, with experimental indices and initial-noise seeds 10000–11023.

Networks run in FP16 at batch size one with guidance scale 3.5. SD3 and SD3.5 use classifier-free guidance, whereas FLUX uses its native guidance embedding. Maximum T5 sequence lengths are 256 for SD3/SD3.5 and 512 for FLUX. Euler uses each model's native timestep schedule, including its resolution-dependent scheduling. Main, anytime, and ablation quality experiments run on RTX PRO 6000; the 1024×1024 study and controlled timings run on A100. Pairing is performed within each experiment and platform.

The implementation uses Python 3.12, PyTorch 2.14.0, Diffusers 0.39.0, and Transformers 4.57.6.

FixUp-FM reuses the completed Euler trace. Restart and SDEdit receive the identical final Euler latent and conditioning, without decoding and re-encoding that latent. Direct Euler, DPM-adaptive, DOPRI5, and SRDS start from the shared initial noise.

Numerical implementation. Reference and corrected states and cached velocities are stored in FP16. Euler and Heun state updates are computed in FP32 and cast back to FP16. For propagation, the implementation uses the commonly scaled residuals $(t_{j+1} - t_{j-1})q_j$ and $(t_{j+1} - t_{j-1})r_j$, whose scale cancels in the rank-one update in exact arithmetic. Residual construction uses mixed precision: reference differences are formed in FP16, time scaling is computed in FP32, and the scaled terms and their subtraction are rounded in FP16 before caching in FP32. Displacements are formed in FP16; inner products, squared norms, and division use FP32. Velocity increments are computed in FP32 and cast to FP16 before addition to the reference velocities. All evaluated interior chords were nonzero, so the zero-chord fallback was never invoked.

I.1 METRICS AND STATISTICAL REPORTING

We use ImageReward-v1.0 with its supplied preprocessing and report the raw prompt-conditioned reward. CLIP denotes the cosine similarity between image and text embeddings from PE-Core-bigG-14-448, multiplied by 100. PE scoring uses the supplied 448×448 image preprocessing and 72-token text context, with FP32 weights and FP16 autocast at batch size one. Both metrics are averaged over prompts, and higher scores are better.

For the ablations, we compute pointwise 95% Student-t confidence intervals from paired, per-prompt score differences. The 1024^2 comparison reports paired mean differences alongside both absolute scores. Differences are computed before rounding.

I.2 CORRECTION BUDGETS AND DISPLAYED OPERATING POINTS

Every integer K from zero through the limit in Table 3 is evaluated independently from the same reference trace. $K = 0$ is the unmodified Euler output.

Direct Euler is matched to FixUp-FM's actual total NFE separately for each prompt before scores are averaged. Curves report mean actual NFE. For $K \geq 2$, the added cost is $K - 1$; for $K = 1$, it is zero when the selected interval is nonterminal and one for a terminal-only correction. A field evaluation includes the model's guidance operation; parallel evaluations of different states count individually. Adaptive initialization, rejected steps, and terminal calls are included.

For each model and base budget $a = 8, 12, \ldots, 32$, we independently select the tested positive correction count that maximizes mean paired IR improvement over samplewise NFE-matched Euler on the 1,024-prompt validation set. Exact ties are resolved by lower mean actual NFE and then smaller K. The resulting $K(a)$ is shared by all 5,000 evaluation prompts and both metrics. SDEdit matches the added budget with a minimum of one evaluation; Restart uses the smallest positive odd budget not below it.

Table 3: Independent correction grid and displayed correction counts. For each base budget, every integer from zero to the maximum K is tested on all three models. The lower rows give the K used in the main figure and table; $a = 4$ is included only in the full sweep.

Base Euler budget a	4	8	12	16	20	24	28	32
Maximum tested K	3	3	3	5	5	8	8	8
Displayed K in the main comparison								
SD3 Medium	—	1	1	1	2	4	2	2
SD3.5 Large	—	1	1	1	2	4	7	8
FLUX.1-dev	—	2	3	3	4	6	8	8

J BASELINE IMPLEMENTATIONS

Flow-model interfaces. The implementation uses descending noise time $\tau = 1 - t \in [0, 1]$, with $x_\tau = (1 - \tau)x_{\text{data}} + \tau z$ and guided velocity $v_\theta(x, \tau) = -f(x, 1 - \tau)$. DPM-adaptive uses the denoiser $D(x, \tau) = x - \tau v_\theta(x, \tau)$, with noise argument $\sigma = \tau$, so its ODE derivative $(x - D)/\tau$ is the original flow velocity. Restart and SDEdit operate in variance-exploding coordinates

$$s = \frac{\tau}{1-\tau}, \qquad y = \frac{x}{1-\tau}, \qquad D_{\text{VE}}(y, s) = x - \tau v_\theta(x, \tau), \qquad x = \frac{y}{1+s}, \quad \tau = \frac{s}{1+s}.$$

The pretrained network and conditioning are unchanged. Solver state precision is retained as described below, while network queries use FP16.

Restart and SDEdit. Restart uses one terminal re-noising cycle, a $\rho = 7$ noise grid, Heun updates, and an Euler update to zero noise. We use noise multiplier 1 and churn $\gamma = 0.05$ for noise levels in $[0.01, 1]$, with FP32 grid construction and FP64 solver states. A grid of m nodes, including both endpoints, costs $2m - 3$ new evaluations. SDEdit uses one repeat of its VE Algorithm 2 with FP64 states, including the stochastic final update. For q evaluations, the grid is $s_0 = 0$ and $s_j = 0.01(s_{\text{start}}/0.01)^{j/q}$ for $j = 1, \ldots, q$.

For each model and method, we evaluate $s_{\text{start}} \in \{0.1, 0.3, 1, 3\}$ on 64 held-out prompts, average IR equally over the seven calibration base/budget pairs, and select one strength. The selected strengths are 0.1 for SDEdit on all models, and 0.1/0.1/0.3 for Restart on SD3/SD3.5/FLUX. These values remain fixed across base budgets, added budgets, and continuation requests. Added-noise generators use seed $10{,}000{,}000 + i$ for prompt index i; independent runs reset this seed, while continuation retains the generator state.

Table 4: Added-NFE budgets used to select the fixed noise strengths. Columns identify the base Euler budget a. Strengths chosen on this grid are held fixed in all subsequent evaluations; displayed comparison budgets are listed in Table 1.

Method	$a = 8$	$a = 12$	$a = 16$	$a = 20$	$a = 24$	$a = 28$	$a = 32$
SD3 Medium							
SDEdit	1	1	1	1	1	1	3
Restart	1	1	1	1	1	1	3
SD3.5 Large							
SDEdit	1	1	1	1	1	2	4
Restart	1	1	1	1	1	3	3
FLUX.1-dev							
SDEdit	1	2	4	4	7	7	7
Restart	1	3	3	3	7	7	7

Adaptive solvers. DPM-adaptive uses the native DPM-Solver-23 controller in `k-diffusion`, order 3, $\eta = 0$, absolute tolerance 0.0078, and FP32 solver states. It integrates from noise time 1 to 10^{-4}, then makes one counted denoising query. The initial log-step is 0.05, PID coefficients are $(0, 1, 0)$, and acceptance safety is 0.81. Relative tolerances are $12.8, 3.2, 0.8, 0.4, 0.2$ for SD3 and FLUX, and $12.8, 3.2, 1.6, 0.8, 0.4$ for SD3.5. The tolerance range is calibrated for evaluation cost; no hard NFE cap is used. DOPRI5 uses SciPy RK45 directly on the flow ODE from $\tau = 1$ to 0, with $\text{rtol} = \text{atol} \in \{0.1, 0.05, 0.02, 0.005\}$, FP64 solver states, and the same FP16 network interface.

SRDS. We use the published Parareal algorithm with Euler coarse and fine solvers and one complete iteration. For an $m \times m$ configuration, the initialization has m coarse intervals and each coarse interval has m fine substeps. Fine solves are batched across intervals, followed by the causal coarse sweep. We evaluate $m = 4, 5, 6$, giving measured total NFEs of 19, 29, and 41 after counting initialization and exact-query reuse. SRDS is evaluated as a sampler from shared initial noise, with its own coarse initialization and matched-cost direct Euler references.

K MAIN-EXPERIMENT RESULTS

K.1 INDEPENDENT CORRECTION SWEEPS

Tables 5–7 report all 51 (a, K) settings per model on 5,000 prompts at 512×512. The main-range summary uses the 120 configurations with $a = 8, 12, \ldots, 32$ and $K > 0$ across the three models. Each model table groups rows by base budget; FixUp and Euler columns use identical samplewise total NFEs. For SD3 and SD3.5, $K = 1$ adds no NFEs for any sample.

Table 5: SD3 Medium: full independent correction sweep on 5,000 prompts at 512×512. Groups identify the base Euler budget a; each row gives the correction count K and mean total NFE. FixUp-FM and Euler are matched samplewise in total NFE. Both metrics are higher-is-better.

		IR ↑		CLIP ↑	
K	Mean total NFE	FixUp	Euler	FixUp	Euler
$a = 4$					
0	4	-0.9390	-0.9390	23.3061	23.3061
1	4	-0.9308	-0.9390	23.0325	23.3061
2	5	-1.1542	-0.2724	22.5154	25.0623
3	6	-1.1259	0.0845	22.5964	25.7760
$a = 8$					
0	8	0.3831	0.3831	26.2311	26.2311
1	8	0.4541	0.3831	26.1799	26.2311
2	9	0.4221	0.4627	26.1225	26.3441
3	10	0.3913	0.5155	26.0856	26.4159
$a = 12$					
0	12	0.5877	0.5877	26.4958	26.4958
1	12	0.6377	0.5877	26.4768	26.4958
2	13	0.6352	0.6023	26.4835	26.4793
3	14	0.6285	0.6170	26.4812	26.5018
$a = 16$					
0	16	0.6434	0.6434	26.5305	26.5305
1	16	0.6767	0.6434	26.5082	26.5305
2	17	0.6831	0.6520	26.5280	26.5475
3	18	0.6815	0.6593	26.5280	26.5600
4	19	0.6799	0.6684	26.5219	26.5743
5	20	0.6778	0.6763	26.5205	26.5786
$a = 20$					
0	20	0.6763	0.6763	26.5786	26.5786
1	20	0.7020	0.6763	26.5720	26.5786
2	21	0.7110	0.6885	26.5910	26.6012
3	22	0.7117	0.6939	26.5959	26.6188
4	23	0.7117	0.7048	26.5963	26.6147
5	24	0.7110	0.7057	26.5928	26.6212

		IR ↑		CLIP ↑	
K	Mean total NFE	FixUp	Euler	FixUp	Euler
$a = 24$					
0	24	0.7057	0.7057	26.6212	26.6212
1	24	0.7241	0.7057	26.6197	26.6212
2	25	0.7336	0.7076	26.6409	26.6186
3	26	0.7351	0.7138	26.6465	26.6268
4	27	0.7361	0.7149	26.6511	26.6273
5	28	0.7359	0.7225	26.6483	26.6435
6	29	0.7358	0.7223	26.6520	26.6458
7	30	0.7354	0.7290	26.6505	26.6474
8	31	0.7353	0.7355	26.6509	26.6633
$a = 28$					
0	28	0.7225	0.7225	26.6435	26.6435
1	28	0.7365	0.7225	26.6313	26.6435
2	29	0.7455	0.7223	26.6479	26.6458
3	30	0.7473	0.7290	26.6539	26.6474
4	31	0.7482	0.7355	26.6563	26.6633
5	32	0.7484	0.7367	26.6587	26.6747
6	33	0.7491	0.7389	26.6595	26.6804
7	34	0.7492	0.7440	26.6591	26.6874
8	35	0.7490	0.7409	26.6559	26.6864
$a = 32$					
0	32	0.7367	0.7367	26.6747	26.6747
1	32	0.7481	0.7367	26.6704	26.6747
2	33	0.7559	0.7389	26.6803	26.6804
3	34	0.7577	0.7440	26.6809	26.6874
4	35	0.7587	0.7409	26.6842	26.6864
5	36	0.7594	0.7430	26.6851	26.6789
6	37	0.7600	0.7457	26.6872	26.6833
7	38	0.7599	0.7470	26.6873	26.6849
8	39	0.7599	0.7431	26.6852	26.7068

Table 6: SD3.5 Large: full independent correction sweep on 5,000 prompts at 512×512. Groups identify the base Euler budget a; each row gives the correction count K and mean total NFE. FixUp-FM and Euler are matched samplewise in total NFE. Both metrics are higher-is-better.

		IR ↑		CLIP ↑	
K	Mean total NFE	FixUp	Euler	FixUp	Euler
$a = 4$					
0	4	-0.9380	-0.9380	22.8896	22.8896
1	4	-0.9129	-0.9380	22.7539	22.8896
2	5	-1.1499	-0.3655	22.1609	24.7191
3	6	-1.1126	-0.0491	22.4423	25.6369
$a = 8$					
0	8	0.2741	0.2741	26.3965	26.3965
1	8	0.3410	0.2741	26.3636	26.3965
2	9	0.2906	0.3535	26.2531	26.5734
3	10	0.2659	0.4116	26.2064	26.6552
$a = 12$					
0	12	0.4708	0.4708	26.6965	26.6965
1	12	0.5285	0.4708	26.6767	26.6965
2	13	0.5201	0.4966	26.6615	26.7605
3	14	0.5108	0.5142	26.6471	26.7728
$a = 16$					
0	16	0.5365	0.5365	26.7832	26.7832
1	16	0.5779	0.5365	26.7731	26.7832
2	17	0.5815	0.5478	26.7919	26.7925
3	18	0.5784	0.5531	26.7832	26.7923
4	19	0.5762	0.5583	26.7793	26.7947
5	20	0.5762	0.5594	26.7759	26.7893
$a = 20$					
0	20	0.5594	0.5594	26.7893	26.7893
1	20	0.5900	0.5594	26.7810	26.7893
2	21	0.5997	0.5670	26.7996	26.7985
3	22	0.5998	0.5749	26.7986	26.8119
4	23	0.5993	0.5753	26.7989	26.8223
5	24	0.5999	0.5844	26.7946	26.8241

		IR ↑		CLIP ↑	
K	Mean total NFE	FixUp	Euler	FixUp	Euler
$a = 24$					
0	24	0.5844	0.5844	26.8241	26.8241
1	24	0.6083	0.5844	26.8135	26.8241
2	25	0.6175	0.5866	26.8230	26.8046
3	26	0.6187	0.5949	26.8217	26.8021
4	27	0.6193	0.5952	26.8214	26.8024
5	28	0.6200	0.6014	26.8192	26.8094
6	29	0.6200	0.6057	26.8199	26.8302
7	30	0.6201	0.6021	26.8208	26.8220
8	31	0.6207	0.6037	26.8212	26.8131
$a = 28$					
0	28	0.6014	0.6014	26.8094	26.8094
1	28	0.6197	0.6014	26.8085	26.8094
2	29	0.6283	0.6057	26.8140	26.8302
3	30	0.6313	0.6021	26.8171	26.8220
4	31	0.6326	0.6037	26.8147	26.8131
5	32	0.6329	0.6101	26.8160	26.8142
6	33	0.6333	0.6080	26.8142	26.8173
7	34	0.6333	0.6105	26.8158	26.8183
8	35	0.6340	0.6103	26.8158	26.8166
$a = 32$					
0	32	0.6101	0.6101	26.8142	26.8142
1	32	0.6248	0.6101	26.8136	26.8142
2	33	0.6339	0.6080	26.8198	26.8173
3	34	0.6369	0.6105	26.8191	26.8183
4	35	0.6384	0.6103	26.8194	26.8166
5	36	0.6394	0.6112	26.8217	26.8234
6	37	0.6397	0.6126	26.8199	26.8147
7	38	0.6401	0.6143	26.8212	26.8109
8	39	0.6405	0.6153	26.8225	26.8150

Table 7: FLUX.1-dev: full independent correction sweep on 5,000 prompts at 512×512. Groups identify the base Euler budget a; each row gives the correction count K and mean total NFE. FixUp-FM and Euler are matched samplewise in total NFE. Both metrics are higher-is-better.

		IR ↑		CLIP ↑	
K	Mean total NFE	FixUp	Euler	FixUp	Euler
$a = 4$					
0	4	0.1061	0.1061	25.5192	25.5192
1	4.9872	0.0210	0.4838	25.5006	26.1059
2	5	0.1321	0.4913	25.6644	26.1222
3	6	0.2010	0.6272	25.7766	26.3289
$a = 8$					
0	8	0.7343	0.7343	26.4245	26.4245
1	8.993	0.7602	0.7740	26.5775	26.4611
2	9	0.7747	0.7761	26.6027	26.4665
3	10	0.7821	0.7930	26.6214	26.4632
$a = 12$					
0	12	0.8193	0.8193	26.4627	26.4627
1	12.993	0.8445	0.8351	26.6061	26.4885
2	13	0.8510	0.8360	26.6058	26.4902
3	14	0.8543	0.8383	26.6119	26.4722
$a = 16$					
0	16	0.8488	0.8488	26.4759	26.4759
1	16.9854	0.8679	0.8566	26.5723	26.4824
2	17	0.8718	0.8578	26.5679	26.4842
3	18	0.8734	0.8603	26.5644	26.4720
4	19	0.8754	0.8638	26.5632	26.4577
5	20	0.8777	0.8632	26.5632	26.4392
$a = 20$					
0	20	0.8632	0.8632	26.4392	26.4392
1	20.963	0.8773	0.8673	26.4983	26.4525
2	21	0.8811	0.8682	26.5014	26.4525
3	22	0.8829	0.8676	26.4959	26.4308
4	23	0.8846	0.8695	26.4934	26.4346
5	24	0.8865	0.8765	26.4892	26.4214

		IR ↑		CLIP ↑	
K	Mean total NFE	FixUp	Euler	FixUp	Euler
$a = 24$					
0	24	0.8765	0.8765	26.4214	26.4214
1	24.9176	0.8863	0.8815	26.4569	26.4208
2	25	0.8902	0.8814	26.4609	26.4195
3	26	0.8920	0.8827	26.4568	26.4092
4	27	0.8939	0.8798	26.4523	26.4093
5	28	0.8952	0.8801	26.4504	26.4017
6	29	0.8959	0.8804	26.4486	26.3974
7	30	0.8972	0.8804	26.4455	26.3954
8	31	0.8980	0.8784	26.4452	26.3746
$a = 28$					
0	28	0.8801	0.8801	26.4017	26.4017
1	28.8758	0.8867	0.8798	26.4273	26.4006
2	29	0.8909	0.8804	26.4333	26.3974
3	30	0.8928	0.8804	26.4315	26.3954
4	31	0.8947	0.8784	26.4287	26.3746
5	32	0.8959	0.8764	26.4263	26.3735
6	33	0.8970	0.8792	26.4268	26.3684
7	34	0.8981	0.8793	26.4252	26.3699
8	35	0.8990	0.8799	26.4228	26.3530
$a = 32$					
0	32	0.8764	0.8764	26.3735	26.3735
1	32.843	0.8813	0.8790	26.3907	26.3717
2	33	0.8849	0.8792	26.4032	26.3684
3	34	0.8873	0.8793	26.4031	26.3699
4	35	0.8885	0.8799	26.4011	26.3530
5	36	0.8900	0.8790	26.3996	26.3383
6	37	0.8908	0.8776	26.3981	26.3406
7	38	0.8921	0.8798	26.3967	26.3513
8	39	0.8928	0.8789	26.3965	26.3414

K.2 SAMPLING BASELINE RESULTS

Table 8: SRDS sampling results on 5,000 prompts per model. Euler uses the same initial noise and total NFE. Grid specifies the coarse–fine configuration.

		SRDS		NFE-matched Euler	
Grid	Total NFE	IR ↑	CLIP ↑	IR ↑	CLIP ↑
SD3 Medium					
4×4	19	0.2628	25.7918	0.6684	26.5743
5×5	29	0.4930	26.0984	0.7223	26.6458
6×6	41	0.5815	26.2789	0.7443	26.7012
SD3.5 Large					
4×4	19	-0.2146	25.1255	0.5583	26.7947
5×5	29	0.0042	25.3601	0.6057	26.8302
6×6	41	0.1316	25.6046	0.6192	26.8096
FLUX.1-dev					
4×4	19	0.8404	26.5195	0.8638	26.4577
5×5	29	0.8981	26.5783	0.8804	26.3974
6×6	41	0.9127	26.5631	0.8781	26.3413

Table 9: Complete adaptive-solver sweeps on 5,000 prompts per model. The panels have separate tolerance grids. DPM-adaptive uses absolute tolerance 0.0078, and its NFE includes rejected steps and terminal denoising. DOPRI5 uses $\mathrm{rtol} = \mathrm{atol} = \epsilon$.

(a) DPM-adaptive

rtol	Mean total NFE	IR ↑	CLIP ↑
SD3 Medium			
12.8	22.1392	0.6942	26.5337
3.2	25.1548	0.6994	26.5597
0.8	30.4756	0.7077	26.5864
0.4	34.3444	0.7090	26.5817
0.2	39.2542	0.7126	26.5831
SD3.5 Large			
12.8	22.9234	0.5194	26.5446
3.2	26.104	0.5281	26.5767
1.6	28.6396	0.5439	26.6105
0.8	32.2606	0.5527	26.6337
0.4	36.889	0.5573	26.6488
FLUX.1-dev			
12.8	22.0504	0.8711	26.3381
3.2	25.0126	0.8699	26.3266
0.8	30.0052	0.8683	26.2759
0.4	33.1558	0.8669	26.2702
0.2	37.0582	0.8690	26.2802

(b) DOPRI5

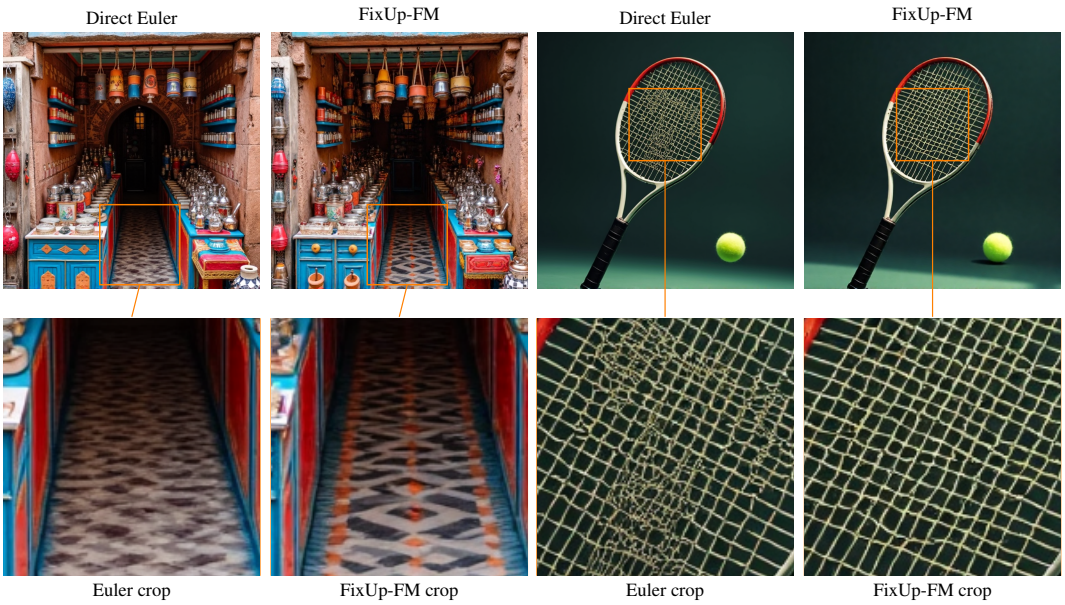
ϵ	Mean total NFE	IR ↑	CLIP ↑
SD3 Medium			
0.1	26	-0.0599	24.4405
0.05	30.6236	0.0287	24.7441
0.02	35.2352	0.1850	25.2226
0.005	47.4692	0.3589	25.7930
SD3.5 Large			
0.1	25.9988	-0.7789	22.4932
0.05	30.014	-0.5350	23.4696
0.02	35.8496	-0.2218	24.5645
0.005	49.7864	0.0870	25.5014
FLUX.1-dev			
0.1	25.9364	0.6975	26.0291
0.05	26.402	0.7327	26.1268
0.02	31.7912	0.7435	26.1444
0.005	45.6512	0.7758	26.2077

L QUALITATIVE COMPARISONS

Figure 6 shows four manually selected FLUX.1-dev examples from the 512×512 experiments.

(a) $a = 32$, $K = 4$, total NFE $= 35$
Prompt: Colorful Moroccan Souk Tea Stall

(b) $a = 20$, $K = 4$, total NFE $= 23$
Prompt: Racket head approaching the ball



(c) $a = 32$, $K = 4$, total NFE $= 35$
Prompt: Microchip chip glowing blue

(d) $a = 24$, $K = 4$, total NFE $= 27$
Prompt: Illustrate a colorful vegetable basket overflowing with vibrant carrots, crisp lettuce, and plump tomatoes, with a few sprigs of fresh rosemary and thyme adding a pop of color to the arrangement.

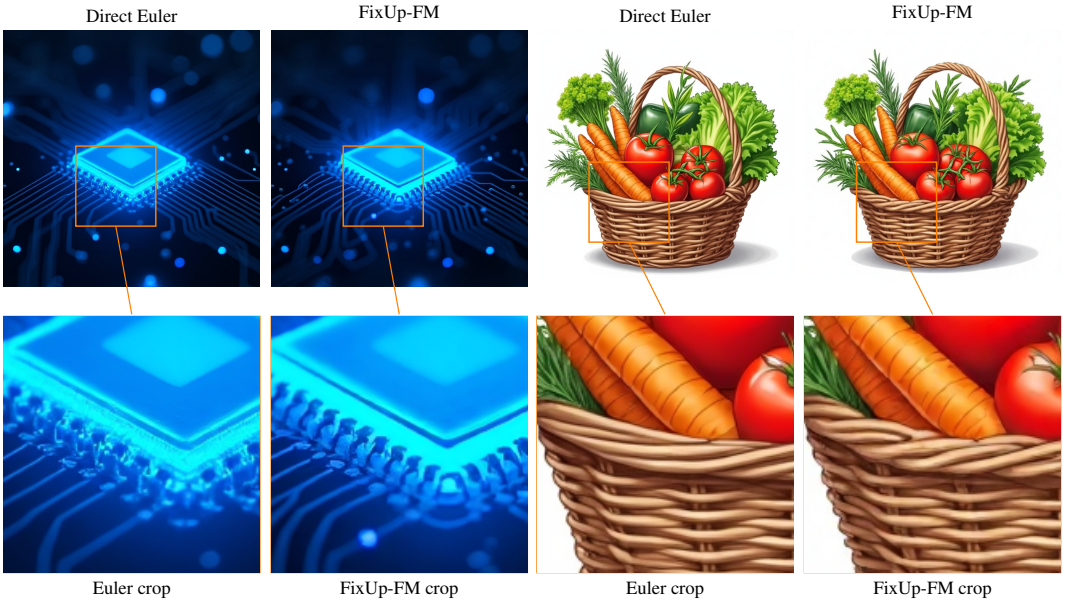


Figure 6: **Qualitative comparisons.** Four selected FLUX.1-dev examples from the 512×512 experiments at equal total NFE. Each pair uses the same prompt and initial noise and shows Direct Euler and FixUp-FM together with enlarged crops of the outlined region, using identical crop coordinates and scale. The base Euler budget a and correction count K refer to FixUp-FM.

M ANYTIME CONTINUATION

All configurations use 3,000 prompts per model and a shared Euler-32 reference. A request of size q is an allowance of q new evaluations. FixUp-FM selects $q+1$ intervals on the first request, exploiting exact endpoint reuse, and q new intervals on each subsequent request. Each pass retains previous corrections and selects only uncorrected intervals, with the reference trace and propagation operators fixed. Cached model evaluations are reused only at the same state and time. At a cumulative allowance of six evaluations, each schedule has corrected seven different intervals.

We evaluate request schedules $+1 \times 6$, $+2 \times 3$, $+3 \times 2$, and a single $+6$ for FixUp-FM and SDEdit. Restart uses $+1 \times 6$, $+3 \times 2$, and a single $+5$. Restart and SDEdit retain their current latent and random-generator state between requests, with the same strengths as in the main comparison. Each schedule starts independently from the common Euler output. Every completed request is scored.

One-shot FixUp entries use a single request from the initial Euler output at each added budget from one to six; repeated-request entries continue from the preceding output. Direct Euler is generated afresh at 32–38 NFEs from the same initial noise. Table columns give actual cumulative added NFE b, with total NFE $32 + b$. Dashes denote budgets not evaluated by that request pattern.

Table 10: SD3 Medium: anytime continuation and one-shot correction from shared Euler-32 outputs on 3,000 prompts. Columns give cumulative added NFE b, with total NFE $32 + b$. Each one-shot entry starts from the initial Euler output; repeated requests build on the preceding output. Dashes denote unevaluated request endpoints.

(a) ImageReward ↑

Method / request pattern	+0	+1	+2	+3	+4	+5	+6
Direct Euler	0.7231	0.7282	0.7306	0.7291	0.7308	0.7329	0.7355
FixUp, one-shot	0.7231	0.7423	0.7444	0.7452	0.7461	0.7466	0.7468
FixUp, repeated +1	0.7231	0.7423	0.7457	0.7472	0.7470	0.7468	0.7466
FixUp, repeated +2	0.7231	—	0.7444	—	0.7469	—	0.7471
FixUp, repeated +3	0.7231	—	—	0.7452	—	—	0.7469
SDEdit, one-shot	0.7231	0.7304	0.7320	0.7321	—	—	0.7315
SDEdit, repeated +1	0.7231	0.7304	0.7323	0.7316	0.7283	0.7266	0.7189
SDEdit, repeated +2	0.7231	—	0.7320	—	0.7337	—	0.7311
SDEdit, repeated +3	0.7231	—	—	0.7321	—	—	0.7343
Restart, one-shot	0.7231	0.7126	—	0.7199	—	0.7263	—
Restart, repeated +1	0.7231	0.7126	0.7036	0.6926	0.6799	0.6695	0.6548
Restart, repeated +3	0.7231	—	—	0.7199	—	—	0.7163

(b) CLIP ↑

Method / request pattern	+0	+1	+2	+3	+4	+5	+6
Direct Euler	26.592	26.614	26.626	26.618	26.609	26.610	26.614
FixUp, one-shot	26.592	26.599	26.599	26.603	26.602	26.604	26.603
FixUp, repeated +1	26.592	26.599	26.602	26.604	26.603	26.602	26.598
FixUp, repeated +2	26.592	—	26.599	—	26.605	—	26.602
FixUp, repeated +3	26.592	—	—	26.603	—	—	26.603
SDEdit, one-shot	26.592	26.587	26.579	26.583	—	—	26.574
SDEdit, repeated +1	26.592	26.587	26.604	26.603	26.599	26.580	26.565
SDEdit, repeated +2	26.592	—	26.579	—	26.598	—	26.600
SDEdit, repeated +3	26.592	—	—	26.583	—	—	26.590
Restart, one-shot	26.592	26.575	—	26.567	—	26.565	—
Restart, repeated +1	26.592	26.575	26.564	26.548	26.542	26.534	26.510
Restart, repeated +3	26.592	—	—	26.567	—	—	26.573

Table 11: SD3.5 Large: anytime continuation and one-shot correction from shared Euler-32 outputs on 3,000 prompts. Columns give cumulative added NFE b, with total NFE $32 + b$. Each one-shot entry starts from the initial Euler output; repeated requests build on the preceding output. Dashes denote unevaluated request endpoints.

(a) ImageReward ↑

Method / request pattern	+0	+1	+2	+3	+4	+5	+6
Direct Euler	0.5958	0.5943	0.5941	0.5981	0.5993	0.6020	0.6051
FixUp, one-shot	0.5958	0.6205	0.6241	0.6256	0.6266	0.6269	0.6275
FixUp, repeated +1	0.5958	0.6205	0.6265	0.6272	0.6261	0.6244	0.6229
FixUp, repeated +2	0.5958	—	0.6241	—	0.6278	—	0.6259
FixUp, repeated +3	0.5958	—	—	0.6256	—	—	0.6273
SDEdit, one-shot	0.5958	0.6039	0.6029	0.6027	—	—	0.6004
SDEdit, repeated +1	0.5958	0.6039	0.6037	0.6003	0.5957	0.5907	0.5826
SDEdit, repeated +2	0.5958	—	0.6029	—	0.6026	—	0.6000
SDEdit, repeated +3	0.5958	—	—	0.6027	—	—	0.6030
Restart, one-shot	0.5958	0.5777	—	0.5959	—	0.5992	—
Restart, repeated +1	0.5958	0.5777	0.5606	0.5447	0.5250	0.5078	0.4898
Restart, repeated +3	0.5958	—	—	0.5959	—	—	0.5954

(b) CLIP ↑

Method / request pattern	+0	+1	+2	+3	+4	+5	+6
Direct Euler	26.767	26.775	26.772	26.768	26.777	26.776	26.778
FixUp, one-shot	26.767	26.777	26.773	26.776	26.778	26.776	26.778
FixUp, repeated +1	26.767	26.777	26.777	26.781	26.775	26.772	26.771
FixUp, repeated +2	26.767	—	26.773	—	26.778	—	26.777
FixUp, repeated +3	26.767	—	—	26.776	—	—	26.784
SDEdit, one-shot	26.767	26.792	26.786	26.770	—	—	26.757
SDEdit, repeated +1	26.767	26.792	26.779	26.772	26.771	26.758	26.742
SDEdit, repeated +2	26.767	—	26.786	—	26.774	—	26.780
SDEdit, repeated +3	26.767	—	—	26.770	—	—	26.767
Restart, one-shot	26.767	26.759	—	26.760	—	26.765	—
Restart, repeated +1	26.767	26.759	26.750	26.740	26.730	26.705	26.675
Restart, repeated +3	26.767	—	—	26.760	—	—	26.748

Table 12: FLUX.1-dev: anytime continuation and one-shot correction from shared Euler-32 outputs on 3,000 prompts. Columns give cumulative added NFE b, with total NFE $32 + b$. Each one-shot entry starts from the initial Euler output; repeated requests build on the preceding output. Dashes denote unevaluated request endpoints.

(a) ImageReward ↑

Method / request pattern	+0	+1	+2	+3	+4	+5	+6
Direct Euler	0.8693	0.8703	0.8723	0.8702	0.8689	0.8679	0.8703
FixUp, one-shot	0.8693	0.8777	0.8797	0.8811	0.8826	0.8835	0.8845
FixUp, repeated +1	0.8693	0.8777	0.8815	0.8851	0.8886	0.8920	0.8944
FixUp, repeated +2	0.8693	—	0.8797	—	0.8861	—	0.8915
FixUp, repeated +3	0.8693	—	—	0.8811	—	—	0.8890
SDEdit, one-shot	0.8693	0.8880	0.8902	0.8889	—	—	0.8830
SDEdit, repeated +1	0.8693	0.8880	0.8893	0.8875	0.8870	0.8841	0.8825
SDEdit, repeated +2	0.8693	—	0.8902	—	0.8931	—	0.8948
SDEdit, repeated +3	0.8693	—	—	0.8889	—	—	0.8927
Restart, one-shot	0.8693	0.8594	—	0.8812	—	0.8982	—
Restart, repeated +1	0.8693	0.8594	0.8494	0.8368	0.8304	0.8252	0.8204
Restart, repeated +3	0.8693	—	—	0.8812	—	—	0.8722

(b) CLIP ↑

Method / request pattern	+0	+1	+2	+3	+4	+5	+6
Direct Euler	26.325	26.323	26.323	26.297	26.284	26.291	26.300
FixUp, one-shot	26.325	26.356	26.356	26.354	26.353	26.350	26.346
FixUp, repeated +1	26.325	26.356	26.362	26.370	26.378	26.389	26.397
FixUp, repeated +2	26.325	—	26.356	—	26.365	—	26.379
FixUp, repeated +3	26.325	—	—	26.354	—	—	26.367
SDEdit, one-shot	26.325	26.568	26.515	26.452	—	—	26.377
SDEdit, repeated +1	26.325	26.568	26.592	26.601	26.588	26.579	26.578
SDEdit, repeated +2	26.325	—	26.515	—	26.541	—	26.540
SDEdit, repeated +3	26.325	—	—	26.452	—	—	26.487
Restart, one-shot	26.325	26.341	—	26.618	—	26.604	—
Restart, repeated +1	26.325	26.341	26.293	26.266	26.250	26.222	26.198
Restart, repeated +3	26.325	—	—	26.618	—	—	26.621

N ABLATION STUDIES

We use SD3.5 Large and 1,000 prompts with indices 256–1255, disjoint from the configuration-screening prompts 0–255. Selection compares the defect-based rule with uniform and random interval selection at $a = 24$, $K = 2, 4, 6$, holding Heun correction and minimum-norm propagation fixed.

Uniform selection rounds the positions $m(a-1)/(K+1)$, $m = 1, \ldots, K$, to the nearest integer, with ties rounded to even. Random selection takes the first K entries of a fixed per-prompt random permutation of $\{0, \ldots, a-1\}$, generated independently of the sampling noise. The same permutation is used across K, giving nested selected sets. Both rules process selected indices in ascending order. All selectors have the same added cost $K - 1$ in these experiments.

Propagation is evaluated at $a = 24, 28, 32$ with fixed $K = 5$, corresponding to four added NFEs at every base budget. Both variants use the same defect-selected intervals and Heun correction rule; the unchanged-velocity control keeps the reference velocities fixed, whereas FixUp-FM uses minimum-norm velocity propagation.

The uncorrected Euler IR/CLIP means on this cohort are 0.5947/26.6912 at $a = 24$, 0.6103/26.7511 at $a = 28$, and 0.6188/26.7682 at $a = 32$.

Table 13: Matched-NFE ablation scores on 1,000 SD3.5 Large prompts. Unchanged denotes reference-velocity reuse. Propagation uses fixed $K = 5$, adding four NFEs at each base budget.

(a) Selection ($a = 24$)

		IR ↑			CLIP ↑		
K	Added NFE	Defect-based	Uniform	Random	Defect-based	Uniform	Random
2	1	0.6306	0.6065	0.6096	26.7085	26.6887	26.6903
4	3	0.6315	0.6159	0.6189	26.7094	26.6950	26.6922
6	5	0.6318	0.6234	0.6226	26.7169	26.6982	26.7039

(b) Propagation

			IR ↑		CLIP ↑	
a	K	Added NFE	Minimum-norm	Unchanged	Minimum-norm	Unchanged
24	5	4	0.6326	0.6271	26.7098	26.6957
28	5	4	0.6439	0.6359	26.7494	26.7380
32	5	4	0.6503	0.6424	26.7741	26.7563

Table 14: Paired FixUp-minus-control differences and pointwise 95% confidence intervals on the same ablation prompts. Positive differences favor FixUp-FM.

			ImageReward		CLIP	
a	K	Control	ΔIR	95% CI	ΔCLIP	95% CI
(a) Selection						
24	2	Uniform	+0.0241	$[+0.0192, +0.0291]$	+0.0198	$[-0.0024, +0.0421]$
24	2	Random	+0.0211	$[+0.0164, +0.0257]$	+0.0182	$[-0.0029, +0.0393]$
24	4	Uniform	+0.0156	$[+0.0105, +0.0207]$	+0.0144	$[-0.0073, +0.0361]$
24	4	Random	+0.0126	$[+0.0074, +0.0177]$	+0.0172	$[-0.0030, +0.0375]$
24	6	Uniform	+0.0084	$[+0.0039, +0.0129]$	+0.0186	$[-0.0008, +0.0380]$
24	6	Random	+0.0092	$[+0.0043, +0.0141]$	+0.0129	$[-0.0063, +0.0321]$
(b) Propagation						
24	5	Unchanged velocities	+0.0055	$[+0.0017, +0.0094]$	+0.0141	$[-0.0027, +0.0309]$
28	5	Unchanged velocities	+0.0080	$[+0.0047, +0.0112]$	+0.0114	$[-0.0032, +0.0261]$
32	5	Unchanged velocities	+0.0079	$[+0.0047, +0.0112]$	+0.0177	$[+0.0031, +0.0324]$

O HIGHER-RESOLUTION GENERATION

The 1024×1024 study uses the same 5,000 FluxPrompting prompts and index-based seeds as the main experiment, on A100. We construct Euler-24 references using the model-native schedule at this resolution, then independently evaluate $K = 0, \ldots, 8$. Direct Euler covers 24–31 NFEs and is paired at each sample's actual cost. For $K = 1$, all SD3/SD3.5 samples use no new evaluations; FLUX uses one for 4,999 samples and zero for one sample, giving mean total NFE 24.9998. Table 15 reports the full sweep.

Table 15: Full 1024×1024 correction sweep from Euler-24 references on 5,000 prompts per model. Euler is matched samplewise to FixUp-FM in total NFE; ΔIR is the paired FixUp-minus-Euler difference.

		IR ↑			CLIP ↑	
K	Mean total NFE	FixUp	Euler	ΔIR	FixUp	Euler
SD3 Medium						
0	24	0.7748	0.7748	+0.0000	26.7864	26.7864
1	24	0.7830	0.7748	+0.0082	26.8558	26.7864
2	25	0.7877	0.7821	+0.0057	26.9157	26.8120
3	26	0.7884	0.7851	+0.0033	26.9209	26.8183
4	27	0.7876	0.7930	-0.0054	26.9273	26.8181
5	28	0.7873	0.7984	-0.0112	26.9303	26.8346
6	29	0.7864	0.8052	-0.0188	26.9274	26.8250
7	30	0.7863	0.8056	-0.0193	26.9244	26.8360
8	31	0.7856	0.8097	-0.0240	26.9248	26.8363
SD3.5 Large						
0	24	0.9924	0.9924	+0.0000	27.5029	27.5029
1	24	1.0023	0.9924	+0.0099	27.5253	27.5029
2	25	1.0076	0.9934	+0.0142	27.5795	27.5058
3	26	1.0082	0.9965	+0.0117	27.5985	27.4906
4	27	1.0079	0.9967	+0.0113	27.6030	27.4937
5	28	1.0075	1.0021	+0.0054	27.6073	27.5040
6	29	1.0075	1.0039	+0.0036	27.6060	27.4900
7	30	1.0070	1.0038	+0.0032	27.6096	27.4894
8	31	1.0068	1.0039	+0.0029	27.6103	27.4909
FLUX.1-dev						
0	24	0.9087	0.9087	+0.0000	26.6495	26.6495
1	24.9998	0.9056	0.9096	-0.0040	26.5963	26.6578
2	25	0.9087	0.9096	-0.0009	26.6061	26.6579
3	26	0.9110	0.9085	+0.0026	26.6122	26.6633
4	27	0.9119	0.9105	+0.0014	26.6147	26.6545
5	28	0.9131	0.9130	+0.0001	26.6180	26.6431
6	29	0.9142	0.9109	+0.0032	26.6185	26.6252
7	30	0.9151	0.9173	-0.0023	26.6228	26.6474
8	31	0.9157	0.9165	-0.0008	26.6226	26.6378

P CONTROLLED RUNTIME MEASUREMENTS

Timings use A100, 512×512, batch size one, and two repetitions per prompt: 256 prompts per model for independent runs and 128 for continuation. Repetitions are averaged within each prompt before taking the dataset mean. End-to-end time includes text encoding, base sampling, correction, and final decoding, but excludes model loading, warm-up, offline scoring, and disk writes. GPU synchronization brackets measurements; runs interrupted by other GPU workloads are repeated.

Table 16 lists the actual measured (a, K) settings and paired Euler costs. For nonzero correction counts at $a = 8$–32, the measured overhead ranges from 0.066% to 0.614%. The $K = 0$ entries are no-correction timing controls.

Table 16: Independent-run A100 timings on 256 prompts per model, with two repetitions at 512×512. FixUp-FM and Euler have identical samplewise total NFE. The time difference is paired; overhead is the relative difference between mean end-to-end times.

			End-to-end time (s)		FixUp minus Euler	
a	K	Mean total NFE	Euler	FixUp	Δt (ms)	Overhead (%)
SD3 Medium						
4	1	4	0.3448	0.3465	1.64	0.476
8	1	8	0.5013	0.5043	3.08	0.614
12	1	12	0.6564	0.6580	1.51	0.230
16	1	16	0.8122	0.8138	1.60	0.197
20	1	20	0.9675	0.9694	1.83	0.189
24	2	25	1.1618	1.1662	4.32	0.372
28	2	29	1.3173	1.3214	4.13	0.313
32	4	35	1.5515	1.5607	9.14	0.589
SD3.5 Large						
4	1	4	0.7181	0.7193	1.19	0.166
8	1	8	1.2465	1.2495	3.00	0.241
12	1	12	1.7746	1.7762	1.69	0.095
16	1	16	2.3031	2.3046	1.52	0.066
20	2	21	2.9634	2.9665	3.09	0.104
24	2	25	3.4913	3.4945	3.23	0.093
28	3	30	4.1519	4.1567	4.73	0.114
32	5	36	4.9442	4.9541	9.92	0.201
FLUX.1-dev						
4	0	4	0.7609	0.7623	1.39	0.183
8	0	8	1.3524	1.3535	1.08	0.080
12	3	14	2.2393	2.2456	6.24	0.279
16	5	20	3.1272	3.1373	10.06	0.322
20	3	22	3.4234	3.4319	8.55	0.250
24	8	31	4.7551	4.7705	15.42	0.324
28	5	32	4.9034	4.9154	11.97	0.244
32	8	39	5.9391	5.9557	16.61	0.280

P.1 CONTINUATION LATENCY

Table 17: A100 continuation timings from Euler-32 references, with 128 prompts per model and two repetitions at 512×512. All times are in seconds and b is cumulative added NFE. Request measures the current request, Upgrade sums completed correction requests, and Total also includes the initial Euler generation and decoding. Each request includes decoding.

		SD3 Medium			SD3.5 Large			FLUX.1-dev		
Round	b	Request	Upgrade	Total	Request	Upgrade	Total	Request	Upgrade	Total
Initial Euler										
0	0	1.4241	0.0000	1.4241	4.4060	0.0000	4.4060	4.8940	0.0000	4.8940
Single $+6$										
1	6	0.2827	0.2827	1.7068	0.8410	0.8410	5.2470	0.9371	0.9371	5.8311
Repeated $+1$										
1	1	0.0822	0.0822	1.5064	0.1754	0.1754	4.5814	0.1917	0.1917	5.0856
2	2	0.0834	0.1657	1.5898	0.1754	0.3508	4.7568	0.1916	0.3832	5.2772
3	3	0.0835	0.2491	1.6733	0.1764	0.5272	4.9332	0.1917	0.5749	5.4689
4	4	0.0831	0.3323	1.7564	0.1770	0.7043	5.1102	0.1915	0.7664	5.6604
5	5	0.0834	0.4156	1.8398	0.1771	0.8814	5.2873	0.1917	0.9581	5.8521
6	6	0.0832	0.4988	1.9229	0.1768	1.0582	5.4642	0.1918	1.1499	6.0439
Repeated $+2$										
1	2	0.1231	0.1231	1.5473	0.3082	0.3082	4.7142	0.3413	0.3413	5.2353
2	4	0.1237	0.2469	1.6710	0.3100	0.6182	5.0242	0.3408	0.6821	5.5761
3	6	0.1235	0.3703	1.7945	0.3103	0.9285	5.3345	0.3409	1.0231	5.9170
Repeated $+3$										
1	3	0.1639	0.1639	1.5880	0.4418	0.4418	4.8478	0.4905	0.4905	5.3845
2	6	0.1635	0.3273	1.7515	0.4433	0.8851	5.2911	0.4902	0.9807	5.8746