ChordEdit: One-Step Low-Energy Transport for Image Editing

Supplementary Material

Contents

A. Full Related Work

GAN-based image editing. Prior to the dominance of diffusion models, generative adversarial networks (GANs) provided strong latent-space controllability. This included latent-space traversal methods [7, 23], encoder-based inversion for real image editing, text-driven manipulation, and CLIP-guided zero-shot domain adaptation. While offering intuitive control, these methods often face challenges in domain generalization and high-resolution reconstruction for real-world images.

Fast One-Step T2I Backbones. The primary enabler for real-time editing is the advent of high-speed generators. These models, often distilled from large-scale diffusion models [12, 21], can synthesize high-fidelity images in a single step. Key examples include adversity-matching or rectified-flow-based generators like SD-Turbo [19], InstaFlow [13], and SwiftBrush-v2 [3]. These backbones provide the foundation for our work, but their fast, non-linear dynamics pose unique challenges for control.

Text-guided editing with diffusion/flow models. Prior work on editing with these backbones falls into several categories with distinct trade-offs:

- **Training-free, Inversion-Required.** This is a common multi-step paradigm. Methods first reconstruct a latent representation of the source image via inversion, then steer the generation process using attention control or guidance, such as in PnP [7, 23], NPI [15, 16], or ProxEdit [6, 8, 11, 18, 24]. While effective, their reliance on iterative inversion and multi-step sampling (e.g., 30-50 steps) makes real-time application infeasible.
- **Training-free, Inversion-Free.** These methods avoid costly per-image inversion, often relying on short trajectories [2, 14]. Differential update strategies, such as FlowEdit [10] and InfEdit [26], are stabilized by sample averaging over multiple steps. However, when forced into the one-step limit, this approach collapses, concentrating high energy and variance into a single, unstable transport step, leading to the failures.
- **Accelerated and Few-Step Editing.** Leveraging fast backbones like SDXL-Turbo [19], methods such as TurboEdit [4] and InstantEdit [1, 5] significantly reduce latency. However, they still require 4-8 steps and often rely on inversion, falling short of true one-step, instant interaction [20, 25].
- **Training-based One-Step Editing.** To achieve true one-step editing, methods like SwiftEdit [9, 17, 22] are proposed. SwiftEdit highlights that one-step editing depends on accurate one-step inversion. It achieves this by training a *dedicated inversion network* to predict the noise, enabling a one-step reconstruction and edit. This reliance

on extra training, however, sacrifices the model-agnostic, training-free nature that is critical for broad applicability. COT Flow [27] uses post-training for one-step transport but is limited to low-resolution 64×64 editing.

Our Method. Different from GAN-based editing and multi-step diffusion editors, **ChordEdit** operates in the challenging training-free, inversion-free, single-step regime. Instead of relying on multi-step differential updates (i.e. InfEdit/ FlowEdit) or trained inversion networks (i.e. SwiftEdit), we introduce a *Chord control field*. This field is constructed directly in the observable residual domain to average and stabilize the high-energy control signal. Paired with a proximal refinement, our method achieves a low-energy, low-variance transport, finally facilitating consistent editing on fast backbones without training or inversion.

B. From Dynamic Optimal Transport to the Chord Control Field

Standing primitives (from the preliminaries). Given prompts c_{src} and c_{tar}, we draw the source and target endpoint images as

$$x_{\mathrm{src}} := x_1 \sim p_1(x \mid c_{\mathrm{src}}), \qquad x_{\mathrm{tar}} := x_0 \sim p_0(x \mid c_{\mathrm{tar}}).$$

Throughout, $x \in \mathbb{R}^d$ denotes the spatial variable and $t \in [0,1]$ is the (diffusion/editing) time, with the convention that $t = 1$ corresponds to the source endpoint and $t = 0$ to the target endpoint.

B.1. Definitions used in the derivation

(D1) Editing density path. We denote by $\{\rho_t\}_{t\in[0,1]}$ a time-indexed family of densities with boundary conditions

$$\rho_1(\cdot) = p_1(\cdot \mid c_{\mathrm{src}}), \qquad \rho_0(\cdot) = p_0(\cdot \mid c_{\mathrm{tar}}). \tag{B.1}$$

(D2) Editing vector field. The *editing vector field* $u_t : \mathbb{R}^d \to \mathbb{R}^d$ is the (complete) probability flow that transports ρ_1 to ρ_0 via the continuity equation

$$\partial_t \rho_t(x) + \nabla \cdot \big(\rho_t(x)\, u_t(x)\big) = 0 \quad \text{for } t \in (0,1). \tag{B.2}$$

We emphasize that u_t is *not* an additive residual on top of another reference field; it is the unique driver of the editing transport in our formulation.

(D3) Observable surrogate of u_t at an anchor. Fix an *anchor* x_τ (in practice we take $x_\tau := x_{\mathrm{src}}$). Let $Q(z,t,c)$ be the model's observable at noisy state z and time t (e.g., noise/velocity prediction). Let $\mathcal{B}_t$ be a time-dependent linear map that converts the model's units to velocity units, and let $K_t(\cdot \mid x_\tau)$ denote the corruption/noising kernel that produces z conditioned on x_τ at time t. Define the observable surrogate

$$\mathbf{R}(x_\tau, t) \;:=\; \mathbb{E}_{z\sim K_t(\cdot|x_\tau)}\big[\, \mathcal{B}_t\big(Q(z,t,c_{\mathrm{tar}}) - Q(z,t,c_{\mathrm{src}})\big)\big]. \tag{B.3}$$

(A1) Local observability (measurement model). At the anchor and within a short temporal window, the surrogate is an unbiased noisy measurement of the editing field:

$$\mathbf{R}(x_\tau, t) \;=\; u_t(x_\tau) \;+\; \varepsilon_t, \qquad \mathbb{E}[\varepsilon_t] = 0. \tag{B.4}$$

(A2) Short-window local homogeneity. Fix a small window $[t-\delta, t]$ with $\delta > 0$. Within the anchor's neighborhood, $u_\xi(x) \approx u$ is (approximately) constant w.r.t. both x and $\xi \in [t-\delta, t]$; the local mass factor $\rho_\xi(x)$ can be absorbed as a (positive) scalar weight.

(A3) Recursive energy prior. The kinetic energy accumulated on $[0, t-\delta]$ induces a quadratic prior around the previous estimate $\hat{u}_{t-\delta}(x_\tau)$ with weight proportional to the elapsed time t.

B.2. Dynamic OT objective with t as progress

The (unregularized) Benamou–Brenier dynamic OT problem directly in the time variable t reads

$$\begin{aligned} \min_{\rho,\, u} \quad & \int_0^1 \int \frac{1}{2} \left\|u_t(x)\right\|^2 \rho_t(x)\, dx\, dt \\ \text{s.t.} \quad & \partial_t \rho_t(x) + \nabla \cdot \big(\rho_t(x)\, u_t(x)\big) = 0, \\ & \rho_1 = p_1(\cdot \mid c_{\mathrm{src}}), \;\; \rho_0 = p_0(\cdot \mid c_{\mathrm{tar}}). \end{aligned} \tag{B.5}$$

This formulation treats u_t as the *complete* field that transports ρ_1 to ρ_0; no reference flow is introduced.

B.3. Local, myopic surrogate of (B.5) and MAP objective

To obtain a causal, single-step estimator of u_t at the anchor, we combine (A1)–(A3) over the short window $[t-\delta, t]$ into the following convex quadratic objective:

$$\Phi_t(u;\, x_\tau) \;=\; t\left\|u - \hat{u}_{t-\delta}(x_\tau)\right\|^2 + \int_{t-\delta}^{t} \left\|u - \mathbf{R}(x_\tau, \xi)\right\|^2 d\xi. \tag{B.6}$$

The first term encodes the recursive energy prior (A3); the second term enforces local agreement with the measurements (A1) under local homogeneity (A2). All global density factors can be absorbed into the (time) weights without changing the closed form below.

B.4. Closed-form minimizer on the window

Differentiating (B.6) w.r.t. u and setting the gradient to zero gives

$$2t\,(u-\hat{u}_{t-\delta})\;+\;2\int_{t-\delta}^{t}(u-\mathbf{R}(x_\tau,\xi))\,d\xi\;=\;0, \quad \text{(B.7)}$$

whence the unique minimizer is

$$u_t^\star(x_\tau)\;=\;\frac{t}{t+\delta}\,\hat{u}_{t-\delta}(x_\tau)\;+\;\frac{1}{t+\delta}\int_{t-\delta}^{t}\mathbf{R}(x_\tau,\xi)\,d\xi. \quad \text{(B.8)}$$

Equation (B.8) is exact under (A1)–(A3).

B.5. Causal first-order approximation (Chord estimator)

For an online single-step implementation, we apply two standard first-order causal approximations on the window $[t-\delta,t]$:

$$\begin{aligned}\int_{t-\delta}^{t}\mathbf{R}(x_\tau,\xi)\,d\xi\;&\approx\;\delta\,\mathbf{R}(x_\tau,t),\\ \hat{u}_{t-\delta}(x_\tau)\;&\approx\;\mathbf{R}(x_\tau,t-\delta).\end{aligned} \quad \text{(B.9)}$$

Substituting (B.9) into (B.8) yields the *Chord* estimate used in our implementation:

$$\boxed{\hat{u}_t(x_\tau)\;=\;\frac{t\,\mathbf{R}(x_\tau,t-\delta)\;+\;\delta\,\mathbf{R}(x_\tau,t)}{t+\delta}.} \quad \text{(B.10)}$$

Remarks on interpretation and accuracy. (i) By construction, u_t (and its estimator $\hat{u}_t$) is the *full* editing flow that drives (B.2) from the source boundary $\rho_1=p_1(\cdot\mid c_{\text{src}})$ to the target boundary $\rho_0=p_0(\cdot\mid c_{\text{tar}})$ in (B.1). (ii) The approximation error of (B.10) relative to (B.8) is $O(\delta)$ under standard smoothness, while measurement noise enters via (B.4). (iii) If desired, one may replace the scalar time-weights (t,δ) in (B.6) by density-weighted effective durations without changing the closed-form structure in (B.8)–(B.10).

C. Unified Comparison-Domain Map and Closed-Form Coefficients

Our framework's model-agnosticism hinges on a linear, time-dependent map $\mathcal{B}_t$, which projects the output Q of any given model into a unified comparison domain U (specifically, the domain of the velocity field u_t). This section provides the closed-form derivations for $\mathcal{B}_t$ under various common model parameterizations.

C.1. General Formulation

We begin with the general forward noising path, which maps a clean image x_0 to a noisy state x_t using a noise sample $\epsilon\sim\mathcal{N}(0,I)$:

$$x_t=\alpha(t)\,x_0+\sigma(t)\,\epsilon. \quad \text{(C.1)}$$

The corresponding continuous-time velocity (or drift) u_t is the time-derivative of this path:

$$u_t:=\dot{x}_t=\dot{\alpha}(t)\,x_0+\dot{\sigma}(t)\,\epsilon. \quad \text{(C.2)}$$

Our goal is to find the map $\mathcal{B}_t$ such that for any model output $\Delta Q=Q(z_t,t,c_{\text{tar}})-Q(z_t,t,c_{\text{src}})$, we have $\mathcal{B}_t(\Delta Q)\approx\Delta u_t$. We compute this difference using a shared-noise sample z_t, which implies a fixed x_t. This "fixed x_t" constraint is key, as it implies $\Delta x_t=0$:

$$\Delta(\alpha(t)\,x_0+\sigma(t)\,\epsilon)=0\implies\alpha(t)\,\Delta x_0+\sigma(t)\,\Delta\epsilon=0. \quad \text{(C.3)}$$

Assuming $\alpha(t)\neq 0$, this provides a direct linear relationship between the change in the predicted data Δx_0 and the change in the predicted noise $\Delta\epsilon$:

$$\Delta x_0=-(\sigma(t)/\alpha(t))\Delta\epsilon. \quad \text{(C.4)}$$

C.2. Noise-Prediction models

This is the most common parameterization, used by models like SD-Turbo. The model directly predicts the noise sample: $Q(z_t,t,c)=\hat{\epsilon}_\theta(z_t,t,c)$. We therefore have $\Delta Q=\Delta\hat{\epsilon}$, and we assume $\Delta\hat{\epsilon}\approx\Delta\epsilon$.

To find the map $\mathcal{B}_t$, we express the velocity difference Δu_t purely in terms of $\Delta\epsilon$ by substituting the constraint for Δx_0:

$$\begin{aligned}\Delta u_t&=\dot{\alpha}(t)\,\Delta x_0+\dot{\sigma}(t)\,\Delta\epsilon && \text{(C.5)}\\ &=\dot{\alpha}(t)\,\Big(-\frac{\sigma(t)}{\alpha(t)}\Delta\epsilon\Big)+\dot{\sigma}(t)\,\Delta\epsilon && \text{(C.6)}\\ &=\Big(\dot{\sigma}(t)-\frac{\dot{\alpha}(t)}{\alpha(t)}\sigma(t)\Big)\Delta\epsilon. && \text{(C.7)}\end{aligned}$$

This gives us a scalar coefficient $A_t^{(\epsilon)}$ that defines the map $\mathcal{B}_t(\Delta Q)=A_t^{(\epsilon)}\Delta Q$:

$$\boxed{A_t^{(\epsilon)}=\dot{\sigma}(t)-\frac{\dot{\alpha}(t)}{\alpha(t)}\,\sigma(t)}. \quad \text{(C.8)}$$

For the common Variance-Preserving (VP) schedule, where $\alpha(t)^2+\sigma(t)^2\equiv 1$, we have $2\alpha\dot{\alpha}+2\sigma\dot{\sigma}=0$, which implies $\dot{\sigma}(t)=-(\alpha(t)/\sigma(t))\dot{\alpha}(t)$. Substituting this into Eq. (C.8) yields a simplified form:

$$\begin{aligned}A_t^{(\epsilon)}&=\Big(-\frac{\alpha(t)}{\sigma(t)}\dot{\alpha}(t)\Big)-\frac{\dot{\alpha}(t)}{\alpha(t)}\sigma(t) && \text{(C.9)}\\ &=-\frac{\dot{\alpha}(t)}{\alpha(t)\sigma(t)}\big(\alpha(t)^2+\sigma(t)^2\big) && \text{(C.10)}\\ &=-\frac{\dot{\alpha}(t)}{\alpha(t)\sigma(t)}. && \text{(C.11)}\end{aligned}$$

$$\boxed{A_t^{(\epsilon)}=-\frac{\dot{\alpha}(t)}{\alpha(t)\,\sigma(t)}}. \quad \text{(C.12)}$$

If the schedule is further parameterized by a continuous-time $\beta(t)$ such that $\alpha(t) = \exp\big(-\frac{1}{2}\int_0^t \beta(s)\,ds\big)$, then $\dot\alpha(t) = -\frac{1}{2}\beta(t)\alpha(t)$. This gives the final form:

$$A_t^{(\epsilon)} = \frac{\beta(t)}{2\,\sigma(t)}. \tag{C.13}$$

C.3. Velocity and Flow-Matching models

This is the most direct case, used by models like InstaFlow. The model output is designed to directly predict the velocity: $Q(z_t, t, c) = u_\theta(z_t, t, c) \approx u_t$.

Therefore, the output difference ΔQ is already in the target comparison domain, and the map $\mathcal{B}_t$ is simply the identity:

$$\boxed{\mathbf{R}(x_\tau, t) = \mathbb{E}_{z\sim K_t(\cdot|x_\tau)}[\Delta u_\theta(z,t)], \qquad \mathcal{B}_t \equiv I}. \tag{C.14}$$

C.4. Other Common Parameterizations

We can derive coefficients for x_0-prediction and v-prediction models using the same principles, assuming a VP schedule for simplicity.

x_0-Prediction Here, $Q = \hat x_0$, so $\Delta Q = \Delta\hat x_0 \approx \Delta x_0$. We map Δx_0 to Δu_t using the constraint $\Delta\epsilon = -(\alpha(t)/\sigma(t))\Delta x_0$:

$$\Delta u_t = \dot\alpha(t)\,\Delta x_0 + \dot\sigma(t)\,\Delta\epsilon \tag{C.15}$$

$$= \dot\alpha(t)\,\Delta x_0 + \dot\sigma(t)\,\Big(-\frac{\alpha(t)}{\sigma(t)}\Delta x_0\Big) \tag{C.16}$$

$$= \Big(\dot\alpha(t) - \frac{\dot\sigma(t)\alpha(t)}{\sigma(t)}\Big)\Delta x_0. \tag{C.17}$$

Using the VP relations $\dot\sigma = -(\alpha/\sigma)\dot\alpha$ and $\alpha^2 + \sigma^2 = 1$:

$$\Delta u_t = \Big(\dot\alpha(t) - \frac{(-\alpha(t)/\sigma(t))\dot\alpha(t)\,\alpha(t)}{\sigma(t)}\Big)\Delta x_0 \tag{C.18}$$

$$= \Big(\dot\alpha(t) + \frac{\alpha(t)^2\dot\alpha(t)}{\sigma(t)^2}\Big)\Delta x_0 \tag{C.19}$$

$$= \dot\alpha(t)\,\Big(\frac{\sigma(t)^2 + \alpha(t)^2}{\sigma(t)^2}\Big)\Delta x_0. \tag{C.20}$$

This yields the map $\mathcal{B}_t(\Delta Q) = A_t^{(x_0)}\Delta Q$:

$$A_t^{(x_0)} = \boxed{\frac{\dot\alpha(t)}{\sigma(t)^2}}. \tag{C.21}$$

v-Prediction Here, $Q = \hat v$, where $v := \alpha\epsilon - \sigma x_0$. Under a VP schedule, the difference Δv relates to $\Delta\epsilon$ as:

$$\Delta v = \alpha\Delta\epsilon - \sigma\Delta x_0 \tag{C.22}$$

$$= \alpha\Delta\epsilon - \sigma\Big(-\frac{\sigma}{\alpha}\Delta\epsilon\Big) \tag{C.23}$$

$$= \Big(\frac{\alpha^2 + \sigma^2}{\alpha}\Big)\Delta\epsilon = \frac{1}{\alpha}\Delta\epsilon. \tag{C.24}$$

Thus, $\Delta\epsilon = \alpha\Delta v$. We map Δv to Δu_t using the coefficient from Eq. (C.12):

$$\Delta u_t = A_t^{(\epsilon)}\Delta\epsilon = \Big(-\frac{\dot\alpha(t)}{\alpha(t)\sigma(t)}\Big)(\alpha(t)\Delta v). \tag{C.25}$$

This gives the map $\mathcal{B}_t(\Delta Q) = A_t^{(v)}\Delta Q$:

$$A_t^{(v)} = \boxed{-\frac{\dot\alpha(t)}{\sigma(t)}}. \tag{C.26}$$

Score-to-drift If we interpret u_t as the drift of the reverse-time SDE associated with the VP schedule, then score-based generative models parameterize this drift by predicting a score field $s_\theta(x_t, t) \approx \nabla_x \log p_t(x_t)$ and converting it to a drift via a scalar time-dependent factor. For a VP forward SDE with noise-rate schedule $\beta(t)$, the reverse dynamics can be written as

$$dx_t = \Big(f_t(x_t) + \beta(t)\,s_\theta(x_t, t)\Big)\,dt + \sqrt{\beta(t)}\,dW_t, \tag{C.27}$$

where f_t collects the parts of the drift that do not depend on the learned score and $\beta(t)$ is the same schedule as in the VP path. For a fixed noisy state x_t the change in drift between two conditioning signals is therefore

$$\Delta u_t = \beta(t)\,\Delta s_\theta(x_t, t). \tag{C.28}$$

Thus these models fit our template with $Q = \hat s$ and $\Delta Q \approx \Delta s_\theta$, and the map $\mathcal{B}_t$ is again time-only:

$$\mathcal{B}_t(\Delta Q) = A_t^{(\text{score})}\,\Delta Q, \qquad A_t^{(\text{score})} = \boxed{\beta(t)}. \tag{C.29}$$

In particular, no dependence on x_t appears in $A_t^{(\text{score})}$ beyond the scalar schedule $\beta(t)$.

Consistency models Consistency models operate on the probability-flow ODE of an underlying diffusion process and learn a "consistency function" $f_\theta(z_t, t)$ that maps any point z_t on an ODE trajectory back to its origin x_0 [21]. Under the parameterization in Eq. (C.1), this means we can simply interpret the model output as an x_0-prediction,

$$Q = \hat x_0 = f_\theta(z_t, t), \qquad \Delta Q \approx \Delta x_0. \tag{C.30}$$

Therefore the same derivation as in the x_0-prediction case applies, and we can reuse the coefficient $A_t^{(x_0)} = \dot\alpha(t)/\sigma(t)^2$:

$$\Delta u_t = A_t^{(x_0)}\,\Delta x_0 = \Big(\frac{\dot\alpha(t)}{\sigma(t)^2}\Big)\Delta x_0. \tag{C.31}$$

Again, the dependence on the particular model enters only through ΔQ; the map $\mathcal{B}_t$ itself remains a linear, time-only scaling.

C.5. Implementation and Numerical Stability

In a practical implementation, the continuous-time derivatives $\dot{\alpha}(t)$ and $\dot{\sigma}(t)$ are approximated using first-order finite differences, e.g., $\dot{\alpha}(t) \approx (\alpha(t) - \alpha(t-\delta))/\delta$.

A critical consideration is that the maps involving $\alpha(t)$ in the denominator (e.g., $A_t^{(\epsilon)}$) become numerically unstable as $\alpha(t) \to 0$, i.e., at $t \approx 1$. Our ChordEdit method queries the field at times t and $t - \delta$ (e.g., $t = 0.90, \delta = 0.15$), which are bounded away from $t = 1$, ensuring $\alpha(t)$ is non-negligible and the map $\mathcal{B}_t$ is well-conditioned.

D. Energy Contraction Property of the Chord Control Field

In this section, we provide the formal justification for the low-energy property of the Chord Control Field. We first demonstrate that the chord estimator, as a form of temporal smoothing, acts as an L^2-energy contraction on the underlying observable proxy field. We then show how this leads to the pointwise energy bound presented in the main text.

We begin by formalizing the chord field $\hat{u}$ as a generalized temporal smoothing operation on the observable proxy field $\mathbf{R}$. The specific estimator derived in Eq. (D.7) is one such example. Let the time-dependent proxy field for a fixed anchor x_τ be denoted $\mathbf{R}(t) := \mathbf{R}(x_\tau, t)$. We define the corresponding chord field $\hat{u}(t) := \hat{u}_t(x_\tau)$ via a causal convolution with a smoothing kernel $K_\delta(s)$:

$$\hat{u}(t) = (K_\delta * \mathbf{R})(t) := \int_{-\infty}^{\infty} K_\delta(s)\mathbf{R}(t-s)\, ds. \quad \text{(D.1)}$$

To match the derivation in Sec. B, this kernel K_δ is assumed to be **non-negative** ($K_\delta(s) \geq 0$), have **unit mass** ($\int K_\delta(s)\, ds = 1$), and be **causal** (e.g., $\mathrm{supp}(K_\delta) \subset [0, \delta]$ for the integral form in Eq. (D.7), or a discrete recursive form). This frames $\hat{u}(t)$ as a weighted average, or expectation, of the recent history of $\mathbf{R}$.

Proposition D.1 (L^2-Energy Contraction)**.** *Let the observable proxy field $\mathbf{R}(t)$ be in $L^2([0,1]; \mathbb{R}^d)$, and let the chord field $\hat{u}(t)$ be generated by convolution with any non-negative, unit-mass kernel K_δ as defined above. The total temporal kinetic energy of the chord field is strictly less than or equal to that of the proxy field:*

$$\int_0^1 \|\hat{u}(t)\|^2\, dt \leq \int_0^1 \|\mathbf{R}(t)\|^2\, dt. \quad \text{(D.2)}$$

Furthermore, the inequality is strict *if K_δ is not a Dirac delta function (i.e., it performs non-trivial averaging) and $\mathbf{R}(t)$ is not almost-everywhere constant on $[0,1]$.*

Proof. The proof relies on the strict convexity of the squared ℓ_2-norm and Jensen's inequality.

1. Pointwise Jensen's Inequality. For any fixed time t, we recognize $\hat{u}(t)$ as the expectation of a vector-valued random variable $Z_s := \mathbf{R}(t-s)$, where the probability measure is $d\mathbb{P}(s) = K_\delta(s)\, ds$.

$$\hat{u}(t) = \int_{\mathbb{R}} \mathbf{R}(t-s)K_\delta(s)\, ds = \mathbb{E}_{s\sim K_\delta}[\mathbf{R}(t-s)]. \quad \text{(D.3)}$$

Let $\varphi(z) = \|z\|^2$. This function is strictly convex on $\mathbb{R}^d$. By Jensen's inequality:

$$\|\hat{u}(t)\|^2 = \varphi(\mathbb{E}[Z_s]) \leq \mathbb{E}[\varphi(Z_s)] = \int_{\mathbb{R}} \|\mathbf{R}(t-s)\|^2 K_\delta(s)\, ds. \quad \text{(D.4)}$$

Equality holds if and only if the random variable Z_s is almost-everywhere constant, i.e., $\mathbf{R}(t-s)$ is constant for s in the support of K_δ.

2. Integration and Fubini's Theorem. We integrate this pointwise inequality over the time interval $t \in [0,1]$. For simplicity, we can consider all functions to be zero-padded outside $[0,1]$ and integrate over $\mathbb{R}$.

$$\begin{aligned}\int_0^1 \|\hat{u}(t)\|^2\, dt &\leq \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \|\mathbf{R}(t-s)\|^2 K_\delta(s)\, ds \right) dt \\ &= \int_{\mathbb{R}} K_\delta(s) \left(\int_{\mathbb{R}} \|\mathbf{R}(t-s)\|^2\, dt \right) ds\end{aligned}$$

where we have exchanged the order of integration by Tonelli's theorem (as the integrand is non-negative). By substituting $\tau = t - s$ (a simple shift), the inner integral becomes $\int_{\mathbb{R}} \|\mathbf{R}(\tau)\|^2\, d\tau = \int_0^1 \|\mathbf{R}(t)\|^2\, dt$.

$$\begin{aligned}\int_0^1 \|\hat{u}(t)\|^2\, dt &\leq \int_{\mathbb{R}} K_\delta(s) \left(\int_0^1 \|\mathbf{R}(t)\|^2\, dt \right) ds \\ &= \left(\int_{\mathbb{R}} K_\delta(s)\, ds \right) \left(\int_0^1 \|\mathbf{R}(t)\|^2\, dt \right).\end{aligned}$$

Since the kernel K_δ has unit mass ($\int K_\delta(s)\, ds = 1$), we arrive at the desired result:

$$\int_0^1 \|\hat{u}(t)\|^2\, dt \leq \int_0^1 \|\mathbf{R}(t)\|^2\, dt. \quad \text{(D.5)}$$

3. Strict Inequality. The inequality is strict if the pointwise Jensen inequality in Eq. (D.4) is strict on a set of t with positive measure. This occurs if $\mathbf{R}(t-s)$ is not constant w.r.t. s on the support of K_δ. If K_δ is not a Dirac delta (i.e., its support has positive measure) and $\mathbf{R}(t)$ is not almost-everywhere constant, this condition will be met, guaranteeing a strict reduction in total energy. □

Remark D.2 (Contraction of Benamou–Brenier Energy)**.** *This proposition extends directly to the full Benamou–Brenier energy functional. The proof above applies pointwise for every $x \in \mathbb{R}^d$.*

$$\int_0^1 \frac{1}{2}\|\hat{u}(x,t)\|^2\, dt \leq \int_0^1 \frac{1}{2}\|\mathbf{R}(x,t)\|^2\, dt. \quad \text{(D.6)}$$

We can then multiply by the non-negative density $\rho_t(x)$ and integrate over x:

$$\mathcal{E}[\hat{u};\rho] = \int_0^1 \int \frac{1}{2}\|\hat{u}(x,t)\|^2 \rho_t(x)\,dx\,dt \leq \int_0^1 \int \frac{1}{2}\|\mathbf{R}(x,t)\|^2 \rho_t(x)\,dx\,dt = \mathcal{E}[\mathbf{R};\rho].$$

Thus, the Chord Control Field $\hat{u}$ generates a dynamic flow with strictly lower (or equal) kinetic energy than the naive proxy field $\mathbf{R}$.

This general L^2-contraction property is instantiated in our specific estimator. The general minimizer from Eq. (D.7) is a convex combination of the prior $\hat{u}_{t-\delta}$ and the observations $\mathbf{R}(\xi)$ over the window.

Corollary D.3 (Pointwise Energy Bound)**.** *The Chord Control Field estimator $\hat{u}_t(x_\tau)$ from the general solution*

$$\hat{u}_t(x_\tau) = \frac{W_t\,\hat{u}_{t-\delta}(x_\tau) + \int_{t-\delta}^{t} \mathbf{R}(x_\tau,\xi)\,d\xi}{W_t + \delta I} \tag{D.7}$$

(assuming W_t is a scalar multiple of identity, $W_t = w_t I$) satisfies the pointwise energy bound:

$$\|\hat{u}_t(x_\tau)\|^2 \leq \frac{w_t\,\|\hat{u}_{t-\delta}(x_\tau)\|^2 + \int_{t-\delta}^{t} \|\mathbf{R}(x_\tau,\xi)\|^2\,d\xi}{w_t + \delta}. \tag{D.8}$$

Furthermore, applying the first-order approximations from the main text (namely $w_t = t$, $\hat{u}_{t-\delta}(x_\tau) \approx \mathbf{R}(x_\tau, t-\delta)$, and $\int_{t-\delta}^{t} \mathbf{R}(\cdot,\xi)d\xi \approx \delta\mathbf{R}(\cdot,t)$) yields the final bound:

$$\|\hat{u}_t(x_\tau)\|^2 \leq \frac{t\,\|\mathbf{R}(x_\tau,t-\delta)\|^2 + \delta\,\|\mathbf{R}(x_\tau,t)\|^2}{t+\delta}. \tag{D.9}$$

Proof. The estimator $\hat{u}_t(x_\tau)$ is a convex combination (a weighted average) of the vectors $\{\hat{u}_{t-\delta}(x_\tau)\} \cup \{\mathbf{R}(x_\tau,\xi)\}_{\xi\in[t-\delta,t]}$. The first inequality follows directly from applying Jensen's inequality to the strictly convex function $\varphi(z) = \|z\|^2$. The second inequality is a direct application of the same principle to the final approximated estimator $\hat{u}_t(x_\tau) = \frac{t}{t+\delta}\mathbf{R}(x_\tau, t-\delta) + \frac{\delta}{t+\delta}\mathbf{R}(x_\tau,t)$, which is a convex combination of the two endpoint proxy fields. □

Lemma D.4 (Local truncation error of explicit Euler under a edit control field)**.** *Consider the ODE over one step $[t_n, t_{n+1}]$ with $t_{n+1} = t_n + h$,*

$$\dot{x}(t) = u(x(t),t), \quad x(t) \in \mathbb{R}^d. \tag{D.10}$$

Assume there exists a set $\mathcal{U} \subset \mathbb{R}^d \times [t_n, t_{n+1}]$ that contains the exact trajectory and on which

$$\|\partial_t u\|_{L^\infty(\mathcal{U})},\ \|\partial_x u\|_{L^\infty(\mathcal{U})},\ \|u\|_{L^\infty(\mathcal{U})} < \infty. \tag{D.11}$$

Let the one–step local truncation error be

$$\tau_{n+1} := x(t_{n+1}) - \big(x(t_n) + h\,f(x(t_n),t_n)\big). \tag{D.12}$$

Then

$$\|\tau_{n+1}\| \leq \frac{h^2}{2}\,M_f, \tag{D.13}$$

where

$$M_f := \sup_{(x,t)\in\mathcal{U}} \big\|\,\partial_t f(x,t) + \partial_x f(x,t)\,f(x,t)\,\big\|. \tag{D.14}$$

In particular, since $f = u$,

$$M_f \leq \|\partial_t u\|_{L^\infty(\mathcal{U})} + \|\partial_x u\|_{L^\infty(\mathcal{U})}\,\|u\|_{L^\infty(\mathcal{U})}. \tag{D.15}$$

Proof. By the variation-of-constants formula,

$$x(t_{n+1}) = x(t_n) + \int_{t_n}^{t_{n+1}} f\big(x(s),s\big)\,ds. \tag{D.16}$$

Using the chain rule and $\dot{x}(s) = f(x(s),s)$,

$$\frac{d}{ds} f\big(x(s),s\big) = \partial_t f\big(x(s),s\big) + \partial_x f\big(x(s),s\big)\,\dot{x}(s) = \partial_t f\big(x(s),s\big) + \partial_x f\big(x(s),s\big)\,f\big(x(s),s\big). \tag{D.17}$$

Integrating (D.17) from t_n to $s \in [t_n, t_{n+1}]$ gives

$$f\big(x(s),s\big) = f\big(x(t_n),t_n\big) + \int_{t_n}^{s} \big(\partial_t f + \partial_x f\,f\big)\big(x(r),r\big)\,dr. \tag{D.18}$$

Insert (D.18) into (D.16) and subtract the explicit Euler update:

$$\tau_{n+1} = \int_{t_n}^{t_{n+1}} \Big(f\big(x(s),s\big) - f\big(x(t_n),t_n\big)\Big)\,ds = \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \big(\partial_t f + \partial_x f\,f\big)\big(x(r),r\big)\,dr\,ds. \tag{D.19}$$

Taking norms and using the definition of M_f,

$$\|\tau_{n+1}\| \leq \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \big\|\big(\partial_t f + \partial_x f\,f\big)\big(x(r),r\big)\big\|\,dr\,ds \leq \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} M_f\,dr\,ds = \frac{h^2}{2}\,M_f, \tag{D.20}$$

which proves (D.13). Since $f = u$, we directly have

$$\big\|\partial_t f + \partial_x f\,f\big\| \leq \|\partial_t u\| + \|\partial_x u\|\,\|u\|, \tag{D.21}$$

from which (D.15) follows by taking the supremum over $\mathcal{U}$. □

Proposition D.5 (Consistency bound for the chord control field)**.** *Let*

$$f_{\mathrm{nai}}(x,t) = \mathbf{R}(x,t), \\ f_{\mathrm{cho}}(x,t) = (K_\delta * \mathbf{R})(x,t), \tag{D.22}$$

where the convolution acts only in time,

$$(K_\delta * \mathbf{R})(x,t) \;=\; \int_{\mathbb{R}} K_\delta(t-s)\,\mathbf{R}(x,s)\,ds, \tag{D.23}$$

with a nonnegative kernel $K_\delta \in L^1(\mathbb{R})$ satisfying $\int_{\mathbb{R}} K_\delta(s)\,ds = 1$. Assume there exists a set $\mathcal{U} \subset \mathbb{R}^d \times [t_n, t_{n+1}]$ that contains the exact trajectory over one step and on which

$$\|\partial_t u\|_{L^\infty(\mathcal{U})},\ \|\partial_x u\|_{L^\infty(\mathcal{U})} \;<\; \infty, \\ \|\partial_t \mathbf{R}\|_{L^\infty(\mathcal{U})},\ \|\partial_x \mathbf{R}\|_{L^\infty(\mathcal{U})},\ \|\mathbf{R}\|_{L^\infty(\mathcal{U})} \;<\; \infty. \tag{D.24}$$

Define the computable consistency proxy for $f = u$ by

$$\mathcal{C}(u;\mathcal{U}) := \|\partial_t u\|_{L^\infty(\mathcal{U})} + \|\partial_x u\|_{L^\infty(\mathcal{U})}\,\|u\|_{L^\infty(\mathcal{U})}. \tag{D.25}$$

Then the chord control field does not increase the consistency bound:

$$\mathcal{C}(K_\delta * \mathbf{R};\mathcal{U}) \;\le\; \mathcal{C}(\mathbf{R};\mathcal{U}). \tag{D.26}$$

Consequently, the local truncation error constant M_f from Lemma D.4 satisfies

$$M_{f_{\mathrm{cho}}} \le \mathcal{C}(K_\delta * \mathbf{R};\mathcal{U}) \le \mathcal{C}(\mathbf{R};\mathcal{U}). \tag{D.27}$$

and hence the explicit Euler local truncation error admits the bound

$$\|\tau_{n+1}^{\mathrm{cho}}\| \;\le\; \frac{h^2}{2}\,\mathcal{C}(K_\delta * \mathbf{R};\,\mathcal{U}) \;\le\; \frac{h^2}{2}\,\mathcal{C}(\mathbf{R};\,\mathcal{U}). \tag{D.28}$$

Proof. Since the convolution acts only in time and K_δ does not depend on x, the spatial and temporal derivatives commute with convolution:

$$\partial_t(K_\delta * \mathbf{R}) \;=\; K_\delta * (\partial_t \mathbf{R}), \qquad \partial_x(K_\delta * \mathbf{R}) \;=\; K_\delta * (\partial_x \mathbf{R}). \tag{D.29}$$

By Young's L^1–L^∞ inequality with $\|K_\delta\|_{L^1} = 1$,

$$\|K_\delta * Z\|_{L^\infty(\mathcal{U})} \;\le\; \|Z\|_{L^\infty(\mathcal{U})}, \qquad Z \in \{\mathbf{R},\,\partial_t\mathbf{R},\,\partial_x\mathbf{R}\}. \tag{D.30}$$

Apply (D.29)–(D.30) termwise in (D.25) with $u = K_\delta * \mathbf{R}$ to obtain

$$\|\partial_t(K_\delta * \mathbf{R})\|_{L^\infty(\mathcal{U})} \le \|\partial_t \mathbf{R}\|_{L^\infty(\mathcal{U})}, \\ \|\partial_x(K_\delta * \mathbf{R})\|_{L^\infty(\mathcal{U})} \le \|\partial_x \mathbf{R}\|_{L^\infty(\mathcal{U})}, \\ \|K_\delta * \mathbf{R}\|_{L^\infty(\mathcal{U})} \le \|\mathbf{R}\|_{L^\infty(\mathcal{U})}. \tag{D.31}$$

Substituting (D.31) into (D.25) yields (D.26). For the connection to the local truncation error, recall from Lemma D.4 that

$$M_f \;=\; \sup_{(x,t)\in\mathcal{U}} \big\|\,\partial_t f(x,t) + \partial_x f(x,t)\,f(x,t)\,\big\| \;\le\; \mathcal{C}(u;\mathcal{U}), \tag{D.32}$$

with $f = u$. Taking $u = \mathbf{R}$ and $u = K_\delta * \mathbf{R}$ gives (D.27), which in turn implies (D.28) via Lemma D.4. □

Theorem D.6 (Global $O(h)$ convergence; chord has smaller constants)**.** *Let $t_n = t_0 + nh$ and consider explicit Euler*

$$x_{n+1} = x_n + h\,u(x_n, t_n). \tag{D.33}$$

where $u \in \{\mathbf{R},\, K_\delta * \mathbf{R}\}$ and K_δ is the time–only kernel from Proposition D.5. Assume there is $\mathcal{U}_T$ containing the exact and numerical trajectories on $[t_0, T]$ such that

$$\|\partial_x u\|_{L^\infty(\mathcal{U}_T)},\ \|\partial_t u\|_{L^\infty(\mathcal{U}_T)}, \\ \|\partial_x \mathbf{R}\|_{L^\infty(\mathcal{U}_T)},\ \|\partial_t \mathbf{R}\|_{L^\infty(\mathcal{U}_T)},\ \|\mathbf{R}\|_{L^\infty(\mathcal{U}_T)} < \infty. \tag{D.34}$$

Define

$$L_u := \|\partial_x u\|_{L^\infty(\mathcal{U}_T)}, \\ M_u := \sup_{(x,t)\in\mathcal{U}_T} \big\|\,\partial_t u + \partial_x u\,u\,\big\|. \tag{D.35}$$

Then for $e_n^u := \|x^u(t_n) - x_n^u\|$ and all $0 \le n \le N$,

$$e_n^u \;\le\; \frac{h\,M_u}{2L_u}\Big(\exp(L_u\,t_n) - 1\Big), \tag{D.36}$$

with the convention $e_n^u \le \frac{h\,t_n}{2}M_u$ when $L_u = 0$. Moreover,

$$L_{f_{\mathrm{cho}}} \le L_{f_{\mathrm{nai}}}, \qquad M_{f_{\mathrm{cho}}} \le M_{f_{\mathrm{nai}}}, \tag{D.37}$$

hence at $t_N = T$ there exist $\mathsf{C}_{\mathrm{cho}} \le \mathsf{C}_{\mathrm{nai}}$ independent of h such that

$$e_N^{\mathrm{cho}} \le \mathsf{C}_{\mathrm{cho}}\,h, \qquad e_N^{\mathrm{nai}} \le \mathsf{C}_{\mathrm{nai}}\,h. \tag{D.38}$$

Proof. Fix u and write $L = \|\partial_x f\|_{L^\infty}$, $M = M_u$. By Lemma D.4,

$$e_{n+1}^u \le (1 + hL)\,e_n^u + \frac{h^2}{2}M. \tag{D.39}$$

Discrete Grönwall gives

$$e_n^u \le \frac{hM}{2}\,\frac{(1+hL)^n - 1}{hL} \le \frac{hM}{2L}\big(\exp(Lt_n) - 1\big), \tag{D.40}$$

which is (D.36). For the comparison, time–only convolution commutes with ∂_t, ∂_x and is L^1–L^∞ nonexpansive, so

$$\|\partial_x(K_\delta * \mathbf{R})\|_\infty \le \|\partial_x \mathbf{R}\|_\infty, \quad \|K_\delta * \mathbf{R}\|_\infty \le \|\mathbf{R}\|_\infty, \\ \|\partial_t(K_\delta * \mathbf{R})\|_\infty \le \|\partial_t \mathbf{R}\|_\infty. \tag{D.41}$$

Hence $L_{f_{\mathrm{cho}}} \le L_{f_{\mathrm{nai}}}$ and, by Proposition D.5, $M_{f_{\mathrm{cho}}} \le M_{f_{\mathrm{nai}}}$, proving (D.37) and (D.38). □

Remark D.7 (Conclusion of Theorem D.6)**.** *The proof establishes the standard $O(h)$ global convergence of the explicit Euler method via the error recurrence (Eq. (D.39)), the local error bound (Lemma D.4), and the discrete Grönwall*

inequality. The key insight is that the global error constant $C(f)$ is smaller for the chord control. This is because the chord field's time derivative $\partial_t f_{\text{cho}}$ exhibits L^∞ contraction (smoothing) via convolution (Proposition D.5), and its field magnitude f_{cho} does not increase. This directly reduces the consistency constant, guaranteeing a smaller global error bound for the same step size h.

Theorem D.6 established that the global $O(h)$ convergence error is governed by the consistency constant $C(f)$, and that $C(f_{\text{cho}}) \le C(f_{\text{nai}})$. This implies that for the same step size h, the global error of the chord-controlled path is bounded by that of the naive path, Global Error$^{\text{cho}} \le$ Global Error$^{\text{nai}}$.

Corollary D.8 (BIBO boundedness under $f = u$)**.** *Assume the editing field u has at most linear growth, e.g., $\|u(x,t)\| \le \beta\|x\| + b$ for all (x,t) in the domain of interest. Then the one–step explicit Euler update $x_{n+1} = x_n + h\,u(x_n, t_n)$ satisfies*

$$\|x_{n+1}\| \le (1 + h\beta)\|x_n\| + h\,b. \tag{D.42}$$

Moreover, if only the trivial bound is desired (without growth assumptions),

$$\|x_{n+1}\| \le \|x_n\| + h\,\|u(x_n, t_n)\|. \tag{D.43}$$

E. Gap to the Optimal Control Field

We now analyze the gap between our estimators and the "true" optimal control $u^\star$. We interpret $u^\star$ as the solution that minimizes the Benamou–Brenier energy under the controlled continuity equation:

$$\partial_t \rho_t + \nabla \cdot \big(\rho_t\, u\big) = 0. \tag{E.1}$$

The following theorem frames $\mathbf{R}(t)$ as a noisy observation of $u^\star(t)$ and shows that the Chord estimator $\hat{u}(t)$ acts as a risk-reducing smoother.

Theorem E.1 (Risk Reduction via Kernel Smoothing)**.** *Let the true optimal control $u^\star : \mathbb{R} \to \mathbb{R}^d$ be C^2 (twice continuously differentiable). We observe the proxy field*

$$\mathbf{R}(t) = u^\star(t) + \eta(t), \tag{E.2}$$

where the noise process $\eta(t)$ satisfies:

(N1) Zero-mean: $\mathbb{E}[\eta(t)] = 0$ for all t.

(N2) Uncorrelated: $\mathbb{E}[\eta(s)\eta(r)^\top] = 0$ for almost every $s \ne r$.

(N3) Bounded variance: $\mathbb{E}[\|\eta(t)\|^2] = \sigma^2(t) \le \bar{\sigma}^2 < \infty$.

*Let K be a **non-negative, unit-mass, second-order kernel**, i.e., $\int K(s)ds = 1$, $\int sK(s)ds = 0$. Define the kernel family $K_\delta(s) = \frac{1}{\delta}K(\frac{s}{\delta})$ for a bandwidth $\delta > 0$, and the chord estimator*

$$\hat{u}(t) = (K_\delta * \mathbf{R})(t) = \int K_\delta(s)\mathbf{R}(t-s)\,ds. \tag{E.3}$$

Then, the Mean Squared Error (Risk) of the chord estimator at time t is bounded by:

$$\boxed{\mathbb{E}[\|\hat{u}(t) - u^\star(t)\|^2] \le \underbrace{c_1\delta^4\|u^{\star\prime\prime}\|_\infty^2}_{\text{Bias}^2} + \underbrace{\|K_\delta\|_{L^2}^2\sigma^2(t)}_{\text{Variance}}}, \tag{E.4}$$

where $c_1 = \frac{1}{4}m_2(K)^2$ and $m_2(K) = \int s^2K(s)ds$. In contrast, the risk of the naive estimator $\mathbf{R}(t)$ is $\mathbb{E}[\|\mathbf{R}(t) - u^\star(t)\|^2] = \sigma^2(t)$. For a non-degenerate kernel, choosing an appropriate δ (see Remark E.3) ensures the chord estimator achieves a strictly lower risk.

Remark E.2 (Causal (One-Sided) Kernels)**.** *The assumption $\int sK(s)ds = 0$ (a second-order kernel) requires K to be symmetric, which is non-causal. If we enforce a **causal, non-negative** kernel (e.g., $\text{supp}(K) \subset [0,1]$ as in our derivation), then $m_1(K) = \int sK(s)ds > 0$. The Taylor expansion (Step 2 in the proof) will be dominated by the first-order term, yielding a bias of $O(\delta)$ and a squared bias of $O(\delta^2)$. The risk bound becomes $O(\delta^2) + O(\delta^{-1})\sigma^2(t)$, but the conclusion of risk reduction still holds.*

Proof. We decompose the Mean Squared Error (Risk) into its squared bias and variance components.

$$\begin{aligned}\mathbb{E}[\|\hat{u}(t) - u^\star(t)\|^2] = &\underbrace{\|\mathbb{E}[\hat{u}(t)] - u^\star(t)\|^2}_{\text{Bias}^2} \\ &+ \underbrace{\mathbb{E}[\|\hat{u}(t) - \mathbb{E}[\hat{u}(t)]\|^2]}_{\text{Variance}}.\end{aligned} \tag{A}$$

1. Bound on the Bias Term. By the linearity of expectation and convolution, and using (N1) ($\mathbb{E}[\eta(t)] = 0$), the expected value of the estimator is:

$$\begin{aligned}\mathbb{E}[\hat{u}(t)] &= \mathbb{E}[(K_\delta * (u^\star + \eta))(t)] \\ &= (K_\delta * u^\star)(t) + (K_\delta * \mathbb{E}[\eta])(t) \\ &= (K_\delta * u^\star)(t).\end{aligned} \tag{E.5}$$

The bias is the difference between this expectation and the true value:

$$\text{Bias}(t) = (K_\delta * u^\star)(t) - u^\star(t) = \int K_\delta(s)\big(u^\star(t-s) - u^\star(t)\big)\,ds. \tag{E.6}$$

We apply a second-order Taylor expansion to $u^\star(t-s)$ around $s = 0$:

$$u^\star(t-s) = u^\star(t) - su^{\star\prime}(t) + \frac{s^2}{2}u^{\star\prime\prime}(t - \theta_s s), \quad \theta_s \in (0,1). \tag{E.7}$$

Substituting this into the bias integral:

$$\begin{aligned}\text{Bias}(t) &= \int K_\delta(s)\left(-su^{\star\prime}(t) + \frac{s^2}{2}u^{\star\prime\prime}(\dots)\right)ds \\ &= -u^{\star\prime}(t)\underbrace{\int sK_\delta(s)\,ds}_{=0 \text{ (2nd-order)}} + \frac{1}{2}\int s^2K_\delta(s)u^{\star\prime\prime}(\dots)\,ds.\end{aligned}$$

The first term vanishes due to the second-order kernel assumption. We bound the remainder:

$$\begin{aligned}\|\mathrm{Bias}(t)\| &\le \frac{1}{2}\int s^2 K_\delta(s)\|u^{\star\prime\prime}(t-\theta_s s)\|\,ds \\ &\le \frac{1}{2}\|u^{\star\prime\prime}\|_\infty \int s^2 K_\delta(s)\,ds.\end{aligned}$$

Since $\int s^2 K_\delta(s)ds = \int s^2 \frac{1}{\delta}K(\frac{s}{\delta})ds = \delta^2 \int u^2 K(u)du = \delta^2 m_2(K)$,

$$\|\mathrm{Bias}(t)\| \le \frac{1}{2}m_2(K)\delta^2\|u^{\star\prime\prime}\|_\infty.$$

Squaring this gives the bound on the first term of (A):

$$\|\mathrm{Bias}(t)\|^2 \le \frac{1}{4}m_2(K)^2\delta^4\|u^{\star\prime\prime}\|_\infty^2. \tag{B}$$

2. Bound on the Variance Term. The variance term is the expected norm of the centered estimator $\zeta(t)$:

$$\zeta(t) = \hat{u}(t) - \mathbb{E}[\hat{u}(t)] = (K_\delta * \eta)(t) = \int K_\delta(s)\eta(t-s)\,ds. \tag{E.8}$$

We write the squared norm as an inner product and apply Fubini's theorem:

$$\begin{aligned}\mathrm{Var}(t) &= \mathbb{E}[\langle\zeta(t),\zeta(t)\rangle] \\ &= \mathbb{E}\left[\left\langle\int K_\delta(s)\eta(t-s)ds, \int K_\delta(r)\eta(t-r)dr\right\rangle\right] \\ &= \iint K_\delta(s)K_\delta(r)\mathbb{E}[\langle\eta(t-s),\eta(t-r)\rangle]\,ds\,dr.\end{aligned}$$

By assumption (N2), the noise is uncorrelated, so the cross-terms where $s \neq r$ (or $t-s \neq t-r$) have zero expectation (a.e.). The integral collapses to the diagonal $s = r$:

$$\begin{aligned}\mathrm{Var}(t) &= \int K_\delta(s)^2\mathbb{E}[\|\eta(t-s)\|^2]\,ds \\ &= \int K_\delta(s)^2\sigma^2(t-s)\,ds.\end{aligned}$$

Bounding the local variance by the value at t (or by $\sup\sigma^2$):

$$\mathrm{Var}(t) \le \sigma^2(t)\int K_\delta(s)^2\,ds = \sigma^2(t)\|K_\delta\|_{L^2}^2. \tag{C}$$

3. Combining the Bounds. Substituting (B) and (C) into (A) yields the theorem's risk bound. In contrast, the risk of the naive estimator (which corresponds to $K = \delta_0$, a Dirac delta) is $\mathbb{E}[\|\mathbf{R}(t) - u^\star(t)\|^2] = \mathbb{E}[\|\eta(t)\|^2] = \sigma^2(t)$. Since for any non-degenerate kernel $\|K_\delta\|_{L^2}^2 < \infty$ (and specifically $\|K_\delta\|_{L^2}^2 \propto \delta^{-1} < \infty$ for $\delta > 0$), the chord estimator achieves variance reduction. By choosing δ appropriately to balance the $O(\delta^4)$ bias and the $O(\delta^{-1})\sigma^2$ variance, the total risk is strictly reduced. □

Remark E.3 (Optimal Bandwidth)**.** *The scaling of the L^2 norm is $\|K_\delta\|_{L^2}^2 = \int \frac{1}{\delta^2}K(\frac{s}{\delta})^2 ds = \frac{\|K\|_{L^2}^2}{\delta}$. The risk bound scales as:*

$$\mathrm{Risk}(\delta) \lesssim c_1\delta^4\|u^{\star\prime\prime}\|_\infty^2 + \frac{\|K\|_{L^2}^2}{\delta}\sigma^2(t). \tag{E.9}$$

Minimizing this with respect to δ gives the classic optimal bandwidth for second-order kernel smoothing, $\delta^\star \asymp \left(\frac{\|K\|_{L^2}^2\sigma^2(t)}{\|u^{\star\prime\prime}\|_\infty^2}\right)^{1/5}$.

Theorem E.4 (Gap to Benamou–Brenier Optimal Energy)**.** *Let $u^\star$ be the true, energy-minimizing optimal control in the space of feasible controls $\mathcal{U}$. Let $\mathcal{U}_\delta \subset \mathcal{U}$ be the subspace of controls that are piecewise linear in time (i.e., chords) on a grid of size δ. Let $P_\delta : \mathcal{U} \to \mathcal{U}_\delta$ be the L^2_ρ-orthogonal projection onto this subspace, where the L^2_ρ norm is induced by the Benamou–Brenier energy functional $\mathcal{E}[u;\rho]$. If we identify the idealized chord estimator $\hat{u}$ with this projection, $\hat{u} = P_\delta u^\star$, the energy gap is bounded by:*

$$\mathcal{E}[\hat{u};\rho] - \mathcal{E}[u^\star;\rho] \le \|(I-P_\delta)u^\star\|_\rho \cdot \|u^\star\|_\rho. \tag{E.10}$$

Furthermore, if $u^\star \in H^1$ (i.e., $\partial_t u^\star$ is in L^2_ρ), the approximation error of the projection is $O(\delta)$, leading to a final bound:

$$\mathcal{E}[\hat{u};\rho] - \mathcal{E}[u^\star;\rho] \le C\delta\|\partial_t u^\star\|_\rho\|u^\star\|_\rho = O(\delta). \tag{E.11}$$

Proof. The proof proceeds in three steps: defining the weighted Hilbert space, applying a projection identity, and bounding the projection error using approximation theory.

1. Weighted Hilbert Space and Energy. We define $H := L^2_\rho([0,1]\times\Omega;\mathbb{R}^d)$ as the Hilbert space of vector fields weighted by the density $\rho_t(x)$. The inner product is:

$$\langle a,b\rangle_\rho := \int_0^1\int_\Omega a(x,t)\cdot b(x,t)\rho_t(x)\,dx\,dt, \tag{E.12}$$

with the induced norm $\|a\|_\rho^2 = \langle a,a\rangle_\rho$. The Benamou–Brenier kinetic energy is $\mathcal{E}[u;\rho] = \frac{1}{2}\|u\|_\rho^2$. We define $P_\delta : H \to \mathcal{U}_\delta$ as the orthogonal projection onto the subspace of piecewise linear functions (chords) with respect to this inner product. We analyze the idealized, noiseless estimator $\hat{u} = P_\delta u^\star$.

2. Projection Identity and Upper Bound. For any $u \in H$, we use the algebraic identity $\langle(P_\delta - I)u, (P_\delta + I)u\rangle_\rho = \langle P_\delta u, P_\delta u\rangle_\rho - \langle u,u\rangle_\rho$, which gives:

$$\|P_\delta u\|_\rho^2 - \|u\|_\rho^2 = \langle(P_\delta - I)u, (P_\delta + I)u\rangle_\rho. \tag{E.13}$$

Applying this to $u = u^\star$ and dividing by 2:

$$\begin{aligned}\mathcal{E}[\hat{u};\rho] - \mathcal{E}[u^\star;\rho] &= \frac{1}{2}\langle(P_\delta - I)u^\star, (P_\delta + I)u^\star\rangle_\rho &\text{(E.14)}\\ &\le \frac{1}{2}\|(I-P_\delta)u^\star\|_\rho \cdot \|(I+P_\delta)u^\star\|_\rho &\text{(E.15)}\end{aligned}$$

by the Cauchy–Schwarz inequality. Since P_δ is an orthogonal projection, its operator norm is $\|P_\delta\| \le 1$. Thus, by the triangle inequality:

$$\|(I+P_\delta)u^\star\|_\rho \le \|Iu^\star\|_\rho + \|P_\delta u^\star\|_\rho \le \|u^\star\|_\rho + \|u^\star\|_\rho = 2\|u^\star\|_\rho. \tag{E.16}$$

Substituting this back, we absorb the constant $\frac{1}{2}\cdot 2 = 1$ into the inequality:

$$\mathcal{E}[\hat{u};\rho] - \mathcal{E}[u^\star;\rho] \le \|(I-P_\delta)u^\star\|_\rho \cdot \|u^\star\|_\rho. \tag{E.17}$$

3. Approximation Error of Chord Space ($O(\delta)$). The term $\|(I-P_\delta)u^\star\|_\rho$ is the minimal L^2_ρ-error when approximating $u^\star$ from the subspace $\mathcal{U}_\delta$. This is a standard result in approximation theory (a Jackson-type inequality). For a function $u^\star$ in the Sobolev space H^1 (meaning its first derivative $\partial_t u^\star$ is in L^2_ρ), the error of the best piecewise linear approximation on a grid of size δ is bounded by the first derivative:

$$\|(I-P_\delta)u^\star\|_\rho \le C_{\text{app}}\delta\|\partial_t u^\star\|_\rho. \tag{E.18}$$

This arises from applying the Poincaré–Wirtinger inequality on each sub-interval $[t_k, t_{k+1}]$ and summing the errors. The constant C_{app} depends on the regularity of the grid but not on δ.

Substituting Eq. (E.18) into Eq. (E.17) gives the final $O(\delta)$ bound:

$$\mathcal{E}[\hat{u};\rho] - \mathcal{E}[u^\star;\rho] \le \big(C_{\text{app}}\delta\|\partial_t u^\star\|_\rho\big)\cdot\|u^\star\|_\rho = O(\delta). \tag{E.19}$$

□

Remark E.5 (Tighter Bound for True Orthogonal Projections)**.** *If $\hat{u}$ is* exactly *the L^2_ρ-orthogonal projection $P_\delta u^\star$, the Pythagorean theorem provides a much tighter (and intuitive) result. Since $u^\star = P_\delta u^\star + (I-P_\delta)u^\star$ is an orthogonal decomposition:*

$$\|u^\star\|^2_\rho = \|P_\delta u^\star\|^2_\rho + \|(I-P_\delta)u^\star\|^2_\rho. \tag{E.20}$$

Therefore, the energy gap is:

$$\begin{aligned}\mathcal{E}[\hat{u};\rho] - \mathcal{E}[u^\star;\rho] &= \frac{1}{2}\big(\|P_\delta u^\star\|^2_\rho - \|u^\star\|^2_\rho\big)\\ &= -\frac{1}{2}\|(I-P_\delta)u^\star\|^2_\rho\\ &\le 0.\end{aligned} \tag{E.21}$$

This confirms that the projection onto the chord subspace never *increases the energy. The $O(\delta)$ bound derived in the main proof is a looser upper bound, but has the advantage of also holding for quasi-projections or causal smoothing operators (like our implemented kernel) whose operator norms are bounded by 1.*

Theorem E.6 (One-Step Error and Stability Condition (Euler))**.** *Assume (A2) ($u \in W^{1,\infty}$), and that $\mathbf{R}$, $K_\delta * \mathbf{R}$ have bounded first derivatives in (x,t) on the relevant domain, with $K_\delta \ge 0$ and $\int K_\delta = 1$. The one-step local truncation errors (from Lemma D.4) for the naive and chord controls are bounded by:*

$$\|\tau^{\text{nai}}_{n+1}\| \le \tfrac{1}{2}h^2\, C_{\text{nai}} \tag{E.22}$$

$$\|\tau^{\text{cho}}_{n+1}\| \le \tfrac{1}{2}h^2\, C_{\text{cho}} \tag{E.23}$$

where C_{nai} and C_{cho} are the global consistency constants:

$$\boxed{\begin{aligned} C_{\text{nai}} &:= \sup_{t\in[0,T]}\Big(\|\partial_t u\| + \|\partial_x u\|\,\|\mathbf{R}\|\Big),\\ C_{\text{cho}} &:= \sup_{t\in[0,T]}\Big(\|\partial_t u\| + \|\partial_x u\|\,\|\hat{u}\|\Big).\end{aligned}} \tag{E.24}$$

Due to the L^∞ contraction properties of the kernel K_δ, $C_{\text{cho}} \le C_{\text{nai}}$. Furthermore, the stability condition for the explicit Euler method (e.g., $hL < 1$) depends only on $L = \sup\|\partial_x u\|$ and is identical for both control fields.

Proof. The proof combines the results from Lemma D.4, Proposition D.5, and Theorem D.6.

1. Specialization of Local Error (from Lemma D.4). Lemma D.4 provides a general one-step error bound:

$$\|\tau_{n+1}\| \le \tfrac{1}{2}h^2 \sup_{s\in[t_n,t_{n+1}]}\Big(\|\partial_t f(x(s),s)\| + \|\partial_x f(x(s),s)\|\cdot\|f(x(s),s)\|\Big).$$

For $f = u$, we have $\partial_x f = \partial_x u$ and $\partial_t f = \partial_t u$. Substituting these for $u_{\text{nai}} = \mathbf{R}$ and $u_{\text{cho}} = \hat{u} = K_\delta * \mathbf{R}$ respectively gives:

$$\begin{aligned}\|\tau^{\text{nai}}_{n+1}\| &\le \tfrac{1}{2}h^2 \sup_{s\in[t_n,t_{n+1}]}\Big(\|\partial_t u + \dot{\mathbf{R}}\| + \|\partial_x u\|\cdot\|\mathbf{R}\|\Big)\\ \|\tau^{\text{cho}}_{n+1}\| &\le \tfrac{1}{2}h^2 \sup_{s\in[t_n,t_{n+1}]}\Big(\|\partial_t u + K_\delta * \dot{\mathbf{R}}\| + \|\partial_x u\|\cdot\|\hat{u}\|\Big).\end{aligned} \tag{E.25}$$

Taking the supremum over the full interval $t \in [0,T]$ (instead of just $[t_n, t_{n+1}]$) defines the global constants C_{nai} and C_{cho} as stated in the theorem.

2. Proof of $C_{\text{cho}} \le C_{\text{nai}}$ (from Prop. D.5). The constant $C(f)$ is the sum of a time-derivative term and a field-magnitude term. *(i) Time-derivative term:* As shown in Prop. D.5, L^∞ contraction by the non-negative, unit-mass kernel K_δ ensures:

$$\begin{aligned}\sup_t\|\partial_t u + K_\delta * \dot{\mathbf{R}}\| &\le \sup_t\|\partial_t u\| + \sup_t\|K_\delta * \dot{\mathbf{R}}\|\\ &\le \sup_t\|\partial_t u\| + \sup_t\|\dot{\mathbf{R}}\|.\end{aligned}$$

This is bounded by $\sup_t \|\partial_t u + \dot{\mathbf{R}}\|$.

(ii) Field-magnitude term: Since $\|\hat{u}\|_\infty \le \|\mathbf{R}\|_\infty$ (Prop. D.5), this term is also bounded by $\sup_t \|\mathbf{R}\|$, which provides a conservative bound for the naive term $\sup_t \|\mathbf{R}\|$.

Since both components of C_{cho} are less than or equal to their counterparts in C_{nai}, we have $C_{\text{cho}} \le C_{\text{nai}}$.

Proposition E.7 (Explicit Euler stability is unaffected by the control design)**.** *Consider one explicit Euler step at time t_n for*

$$\dot{x}(t) = u(x(t), t), \tag{E.26}$$

with $u \in \{\mathbf{R},\ K_\delta * \mathbf{R}\}$ and the time–only convolution

$$(K_\delta * \mathbf{R})(x,t) = \int_{\mathbb{R}} K_\delta(t-s)\,\mathbf{R}(x,s)\,ds, \qquad K_\delta \ge 0,\ \int_{\mathbb{R}} K_\delta = 1. \tag{E.27}$$

Let $\mathcal{U} \subset \mathbb{R}^d \times [t_n, t_{n+1}]$ contain the exact one–step trajectory and assume

$$\|\partial_x u\|_{L^\infty(\mathcal{U})},\ \|\partial_x \mathbf{R}\|_{L^\infty(\mathcal{U})} < \infty. \tag{E.28}$$

Define the step–n Jacobian bound

$$\begin{aligned} L_{\text{nai}} &:= \|\partial_x u\|_{L^\infty(\mathcal{U})} + \|\partial_x \mathbf{R}\|_{L^\infty(\mathcal{U})}, \\ L_{\text{cho}} &:= \|\partial_x u\|_{L^\infty(\mathcal{U})} + \|\partial_x (K_\delta * \mathbf{R})\|_{L^\infty(\mathcal{U})}. \end{aligned} \tag{E.29}$$

Then, by time–only convolution and Young's L^1–L^∞ inequality,

$$L_{\text{cho}} \ \le\ L_{\text{nai}}. \tag{E.30}$$

Consequently, any step–size prescription of the form

$$h \ \le\ \phi\big(\|\partial_x u\|_{L^\infty(\mathcal{U})}\big)\,, \tag{E.31}$$

with $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ nonincreasing (e.g. $\phi(L) = \eta/L$ for a chosen $\eta > 0$), admits a weakest–case bound that is not tightened by using the chord control:

$$h^{\text{cho}}_{\max} \ \ge\ h^{\text{nai}}_{\max}. \tag{E.32}$$

In particular, the linearized one–step growth factors satisfy

$$\begin{aligned} \sup_{(x,t)\in\mathcal{U}} \big\|I + h\,\partial_x f^{\text{cho}}(x,t)\big\| \le 1 + h\,L_{\text{cho}} \ \le\ 1 + h\,L_{\text{nai}} \\ \ge \sup_{(x,t)\in\mathcal{U}} \big\|I + h\,\partial_x f^{\text{nai}}(x,t)\big\|, \end{aligned} \tag{E.33}$$

so any stability target expressed as $1 + h\,L \le 1 + \eta$ (or equivalently $h \le \eta/L$) is never harder to meet under the chord control field.

Proof. Since K_δ acts only in time and is independent of x, differentiation commutes with convolution:

$$\partial_x (K_\delta * \mathbf{R}) \ =\ K_\delta * (\partial_x \mathbf{R}). \tag{E.34}$$

With $\|K_\delta\|_{L^1} = 1$, Young's inequality gives

$$\|\partial_x (K_\delta * \mathbf{R})\|_{L^\infty(\mathcal{U})} \ \le\ \|\partial_x \mathbf{R}\|_{L^\infty(\mathcal{U})}. \tag{E.35}$$

Adding $\|\partial_x u\|_{L^\infty(\mathcal{U})}$ to both sides yields (E.30).

For any u, the Jacobian of the explicit Euler one–step map $x \mapsto x + hu(x,t_n)$ is $I + h\,\partial_x u(x,t_n)$, hence

$$\sup_{(x,t)\in\mathcal{U}} \big\|I + h\,\partial_x u(x,t)\big\| \ \le\ 1 + h\,\|\partial_x u\|_{L^\infty(\mathcal{U})}. \tag{E.36}$$

Therefore any nonincreasing step–size rule (E.31) that enforces a desired upper bound on (E.36) becomes no stricter when $\|\partial_x u\|_{L^\infty}$ is replaced by the smaller value L_{cho}. This proves (E.32) and (E.33). □

3. Stability and Global Error (from Prop. E.7 and Thm. D.6). As shown in Prop. E.7, the stability of the Euler method depends on the Jacobian $\partial_x f = \partial_x u$, which is identical for both fields. The stability condition $hL < 1$ (where $L = \sup \|\partial_x u\|$) is therefore unchanged. As proven in Thm. D.6, the global error is bounded by

$$\max_n \|e_n\| \le \frac{e^{LT} - 1}{L} \cdot \frac{h}{2} \cdot C(f).$$

Since $C_{\text{cho}} \le C_{\text{nai}}$, it follows directly that for the same step size h, the global error bound for the chord control is also smaller or equal. □

Remark E.8 (Condition for Strict Inequality)**.** *The inequality $C_{\text{cho}} \le C_{\text{nai}}$ becomes strict, $C_{\text{cho}} < C_{\text{nai}}$, if the kernel K_δ is non-degenerate (not a Dirac delta, $\delta > 0$) and the proxy field's derivative $\dot{\mathbf{R}}$ is not almost-everywhere constant. In this (typical) case, the L^∞ smoothing of the time-derivative term is strict ($\|K_\delta * \dot{\mathbf{R}}\|_\infty < \|\dot{\mathbf{R}}\|_\infty$), leading to a strictly smaller error constant and a tighter global error bound.*

Corollary E.9 (Explicit Error Ratio)**.** *Under the assumptions of Theorem E.6, the ratio of the global error bounds is*

$$\begin{aligned} &\frac{\textit{Global Error}^{\text{cho}}}{\textit{Global Error}^{\text{nai}}} \\ &\le \frac{\sup_t\big(\|\partial_t u\| + \|\partial_x u\|\,\|\hat{u}\|\big)}{\sup_t\big(\|\partial_t u\| + \|\partial_x u\|\,\|\mathbf{R}\|\big)} \ \le\ 1. \end{aligned} \tag{E.37}$$

Equality holds only in the degenerate case where the smoothing has no effect (e.g., $\delta = 0$ or $\dot{\mathbf{R}} \equiv 0$).

F. Additional Ablation Studies

F.1. Additional Analysis of the Chord Control Field

As established in the main paper, the fundamental hypothesis of our work is that the naive editing field (equivalent to ChordEdit with $\delta = 0$) is inherently unstable for one-step

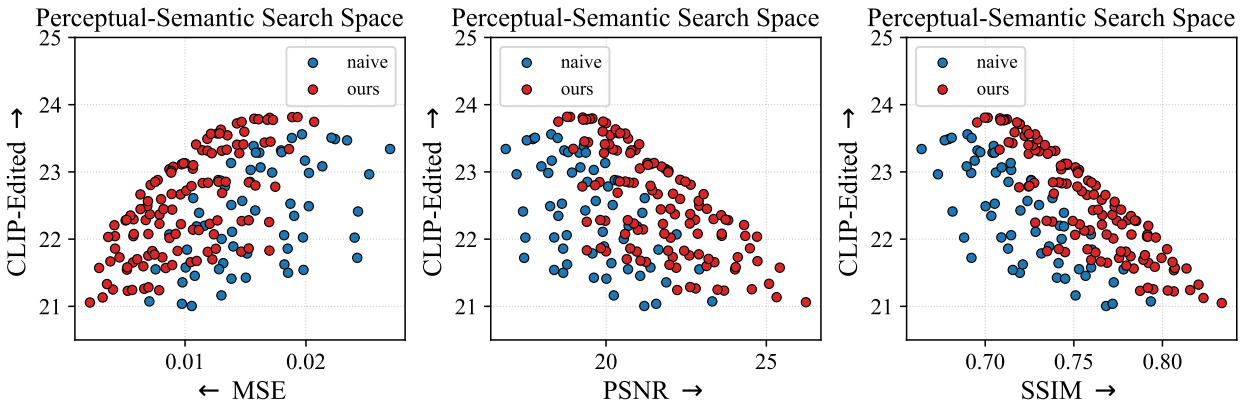

Figure 1. **Raw hyperparameter sweep distribution for CCF analysis.** We compare the naive baseline ($\delta = 0$, blue) against ChordEdit ($\delta \neq 0$, red). The plots visualize the trade-off between semantic alignment (CLIP-Edited score) and three standard background preservation metrics: (left) Mean Squared Error (MSE), (center) Peak Signal-to-Noise Ratio (PSNR), and (right) Structural Similarity Index (SSIM). In all three trade-off spaces, the ChordEdit samples (red) occupy a visibly superior region (e.g., lower MSE, higher PSNR/SSIM for a given CLIP score) than the naive baseline, which exhibits a wider, less stable, and inferior performance distribution.

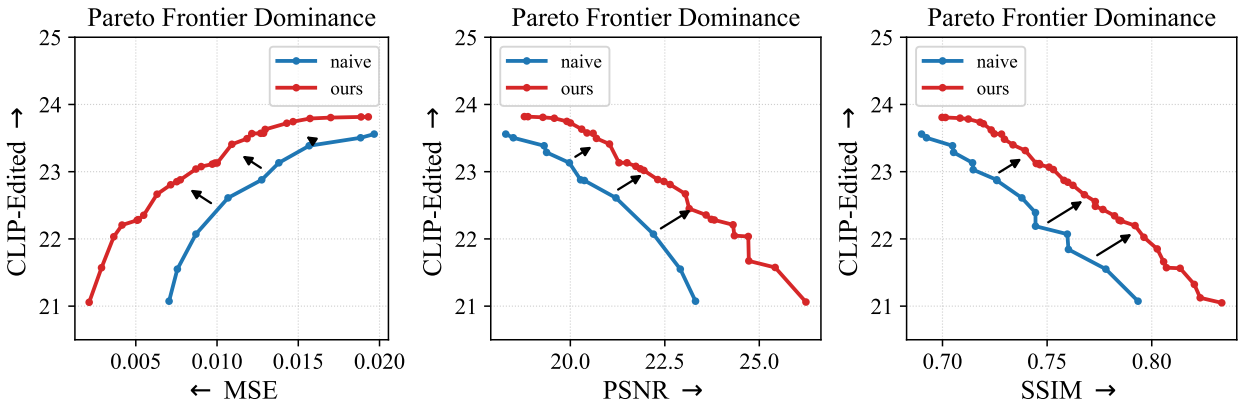

Figure 2. **Pareto dominance of the Chord Control Field.** These frontiers, derived from the data in Figure 1, confirm the strict Pareto dominance of ChordEdit ($\delta \neq 0$, red) over the naive baseline ($\delta = 0$, blue) across all three preservation metrics: (left) MSE vs. CLIP, (center) PSNR vs. CLIP, and (right) SSIM vs. CLIP. In every case, the ChordEdit frontier achieves superior semantic alignment (higher CLIP) for any given level of perceptual fidelity, and vice-versa. This empirically validates that the temporal smoothing induced by $\delta > 0$ is key to resolving the inferior trade-off inherent in the unstable naive approach.

integration. Our Chord Control Field (CCF) resolves this by introducing a temporal smoothing interval $\delta > 0$, which yields a stable, low-energy transport path.

To provide a comprehensive validation of this claim, we expand upon the LPIPS-CLIP Pareto analysis of the main text. We conduct an extensive hyperparameter sweep for both the naive baseline ($\delta = 0$) and ChordEdit ($\delta \neq 0$) and evaluate the trade-off between semantic alignment (CLIP-Edited) and a wider array of background preservation metrics.

Figure 1 visualizes the raw data distributions from this sweep, plotting performance against (left) Mean Squared Error (MSE), (center) Peak Signal-to-Noise Ratio (PSNR), and (right) Structural Similarity Index (SSIM). In all three scatter plots, the ChordEdit samples (red) are visibly concentrated in a superior performance region (e.g., lower MSE, higher PSNR/SSIM for a given CLIP score) compared to the naive baseline (blue), which exhibits a much wider, less stable, and fundamentally inferior distribution.

Figure 2 plots the resulting Pareto frontiers from this data. These results unequivocally demonstrate that ChordEdit strictly Pareto-dominates the naive baseline across all three trade-off spaces. This confirms that the instability of the naive field is not limited to a single perceptual metric (like LPIPS) but is a fundamental flaw. By leveraging the temporally-smoothed, low-energy Chord Control Field, our method consistently achieves a superior and more robust performance envelope, enabling high semantic alignment and high struc-

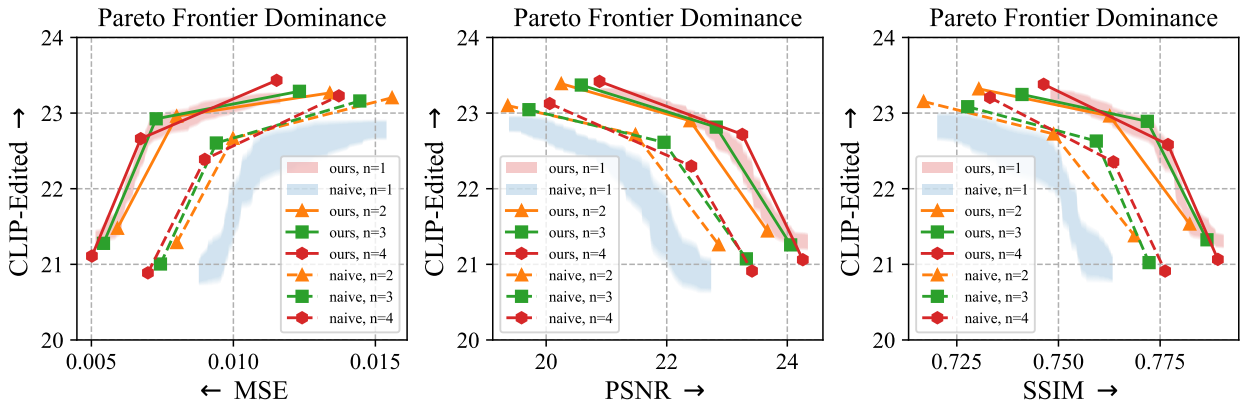

Figure 3. **Noise sample analysis across multiple metrics.** This figure expands the analysis in the main paper by plotting the semantic-preservation trade-off (CLIP-Edited vs. MSE/PSNR/SSIM) as a function of the number of noise samples (n). The overlapping distributions and tight confidence bands confirm that increasing $n > 1$ provides negligible marginal returns. ChordEdit's performance with $n = 1$ is already highly stable and robust, validating our default setting.

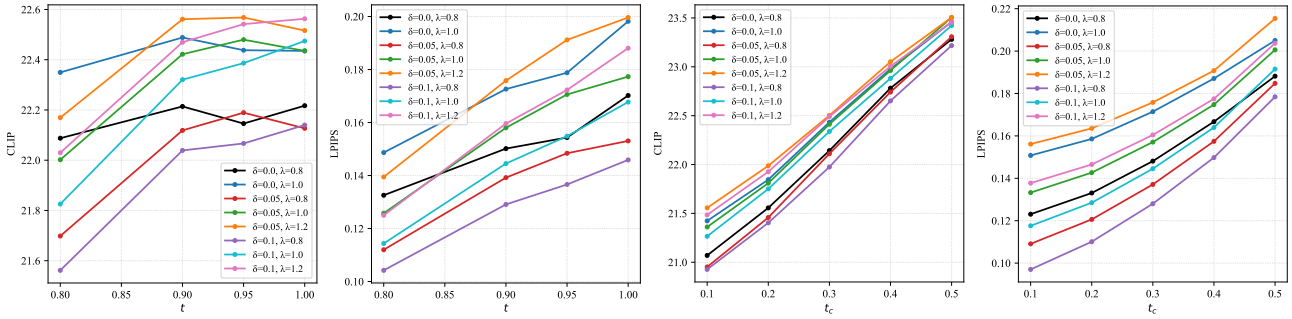

Figure 4. **Ablation study on temporal hyperparameters and step scale.** We analyze the sensitivity of ChordEdit to key parameters. **(Left two panels):** Impact of the main chord time t. Increasing t generally improves semantic alignment (CLIP-Edited ↑) at the cost of background preservation (LPIPS-Unedit ↓). **(Right two panels):** Impact of the proximal refinement time t_c. Increasing t_c robustly improves CLIP score, but also monotonically increases LPIPS distortion. This analysis confirms a clear trade-off space, allowing for the principled selection of our default parameters which balance these competing objectives.

tural preservation simultaneously.

F.2. Additional Analysis of Noise

We validated the robustness of ChordEdit to random noise seeds and established that increasing the number of Monte Carlo (MC) samples (n) yields negligible marginal returns. This finding is attributed to the intrinsically low variance of the Chord Control Field, which achieves stability through temporal smoothing rather than costly MC averaging.

We expand this analysis here. Figure 3 extends the Pareto-frontier analysis from the main text (which used LPIPS-CLIP) to cover a broader range of background preservation metrics: (left) Mean Squared Error (MSE), (center) Peak Signal-to-Noise Ratio (PSNR), and (right) Structural Similarity Index (SSIM).

Consistent with our primary findings, these plots demonstrate that the trade-off between semantic alignment (CLIP-Edited) and structural preservation (MSE/PSNR/SSIM) is virtually independent of the number of noise samples (n). The performance distributions (visualized by the scatter points) and their stability (implied by the confidence bands) are nearly indistinguishable whether using $n = 1$ or multiple samples with our method. This stability is unique to our method; the naive baseline, in contrast, suffers from high intrinsic variance, causing its $n = 1$ performance to be significantly less stable and worse than its multi-sample ($n > 1$) configurations. This strongly confirms our hypothesis: the ChordEdit $n = 1$ configuration is not a compromise but operates robustly at the optimal performance frontier. This result empirically justifies our default use of $n = 1$ for all main experiments, achieving maximum efficiency without sacrificing quality or stability.

F.3. Analysis of Temporal Parameters and Step Scale

In addition to the core analysis of δ in the main paper, the performance of ChordEdit is jointly influenced by several

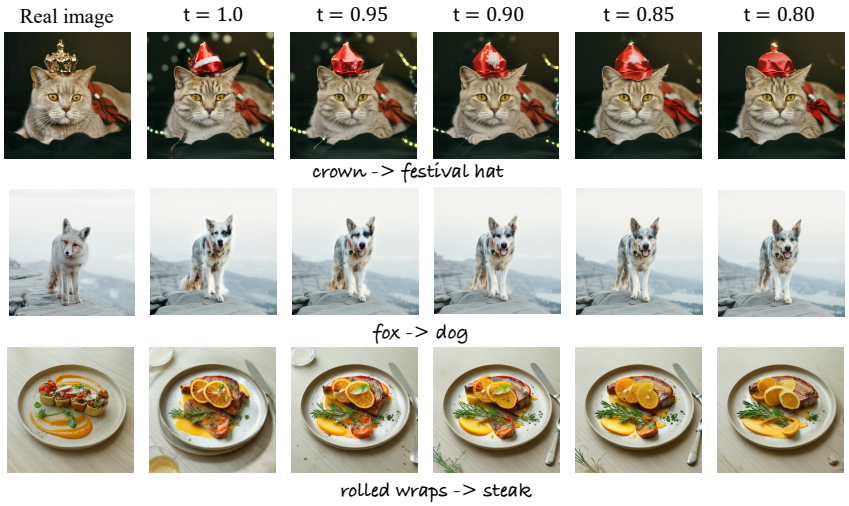

Figure 5. **Qualitative analysis of the main chord time t.**

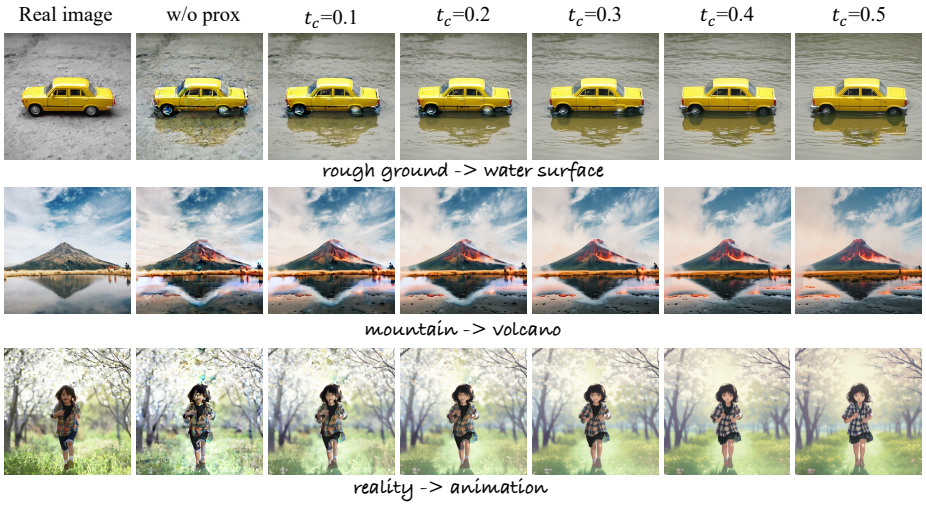

Figure 6. **Qualitative analysis of the proximal refinement time t_c.**

key hyperparameters: the primary chord time t, the step scale λ, and the proximal refinement time t_c. Figure 4 presents a comprehensive ablation study to investigate the sensitivity and trade-offs associated with these parameters.

Analysis of Chord Time t. The left two panels of Figure 4 illustrate the impact of the main chord time t on semantic alignment (CLIP-Edited ↑) and background preservation (LPIPS-Unedit ↓). We observe a distinct trade-off: increasing t (e.g., from 0.80 to 1.00) generally yields stronger semantic alignment, as the model queries the field at a point closer to the final, fully-formed image manifold. However, this comes at the cost of slightly reduced background fidelity (higher LPIPS). The plots also reaffirm our central thesis: the naive baseline ($\delta = 0.0$, black and blue lines) consistently occupies an inferior performance region (lower CLIP for a given LPIPS) compared to our smoothed ChordEdit configurations ($\delta > 0$). This quantitative trade-off is qualitatively visualized in Figure 5, which confirms that $t = 0.90$ provides a robust balance between semantic strength and preservation.

Analysis of Refinement Time t_c. The right two panels of Figure 4 analyze the effect of the proximal refinement time t_c. The results show a strong, monotonic relationship: increasing t_c from 0.1 to 0.5 robustly enhances semantic alignment (CLIP). This confirms the role of the proximal step in "sharpening" the edit to better match the target prior. However, this semantic gain is directly coupled with a degradation in background preservation (rising LPIPS), as

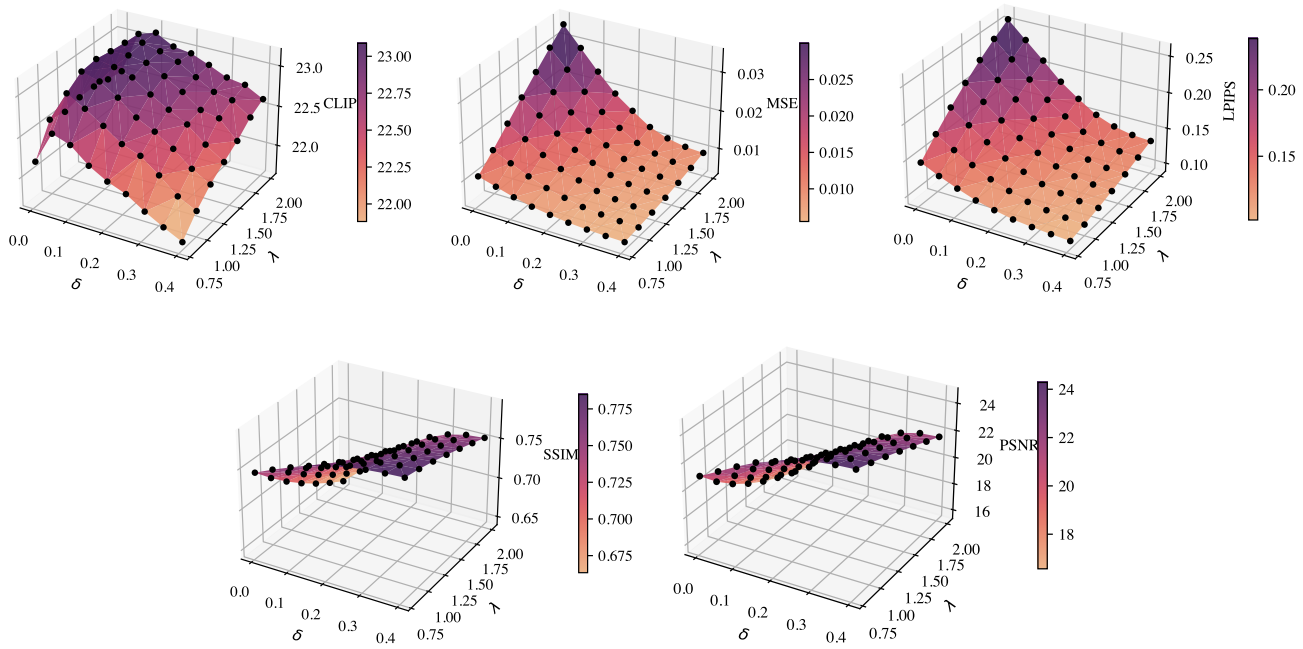

Figure 7. **Quantitative joint-analysis of δ and λ.** These 3D surface plots show the trade-off between semantic alignment (CLIP) and background preservation (LPIPS, MSE, SSIM, PSNR). The naive baseline ($\delta = 0$, the front edge of each plot) is Pareto-inferior, suffering from low CLIP scores and high distortion (high LPIPS/MSE). Increasing δ (our temporal smoothing) robustly and monotonically improves all preservation metrics. Increasing λ (step scale) robustly increases semantic strength (CLIP) at the cost of preservation. Our default parameters are chosen from this smooth, well-behaved trade-off space.

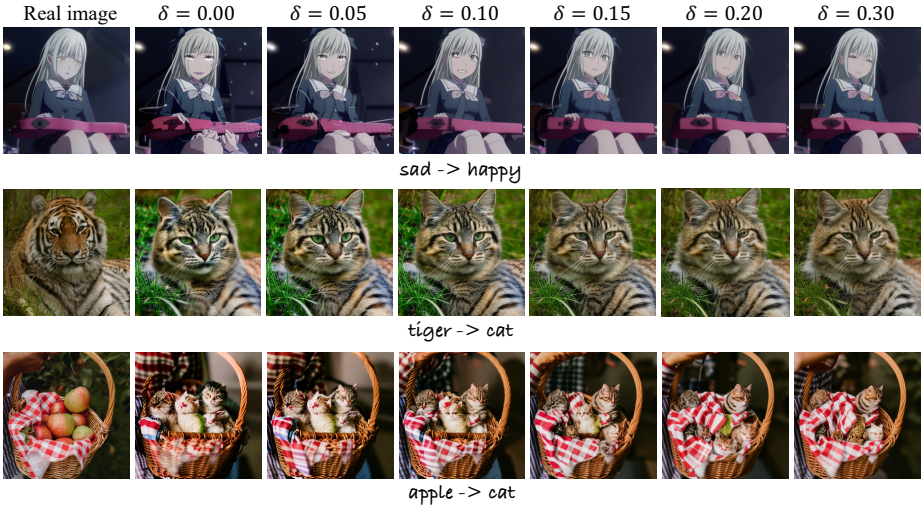

Figure 8. **Qualitative analysis of the temporal window δ.** We fix other parameters and vary δ. The naive baseline ($\delta = 0.00$) consistently fails, producing severe artifacts, distortions, and structural collapse, especially on complex semantic changes (e.g., 'tiger $\rightarrow$ cat' or 'apple $\rightarrow$ cat'). Our Chord Control Field ($\delta > 0$) immediately stabilizes the edit, resolving these failures. A value of $\delta = 0.15$ demonstrates a robust balance between edit stability and semantic strength, validating its choice as our default.

a "stronger" refinement (higher t_c) is more prone to over-editing and affecting non-target regions. Figure 6 provides a clear visual example, showing how increasing t_c strengthens the target semantic at a modest cost to fidelity, justifying our default choice of $t_c = 0.30$.

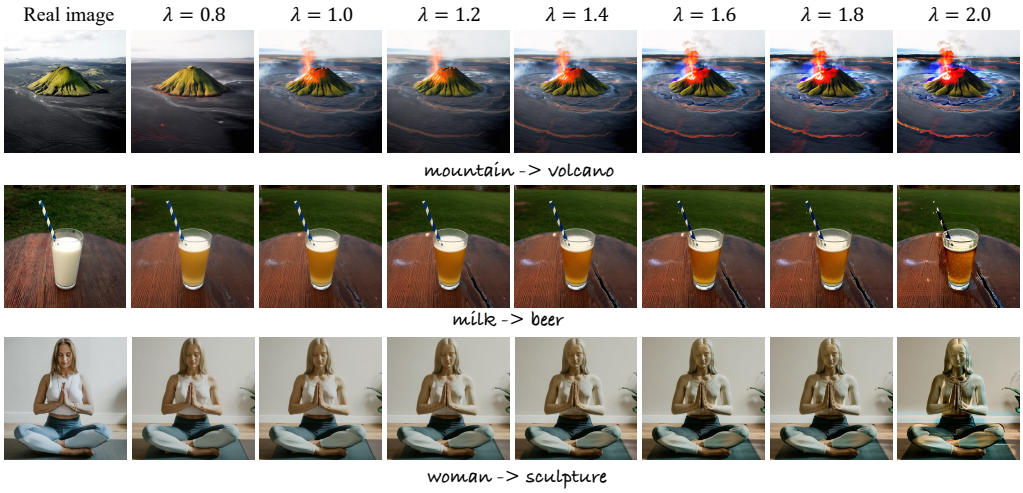

Figure 9. **Qualitative analysis of the step scale λ.** We vary λ while keeping other parameters fixed. λ functions as an intuitive 'edit strength' controller. Small values (e.g., $\lambda = 0.8$) result in subtle, under-edited images ('mountain'). As λ increases, the intensity of the target semantic ('volcano') becomes progressively stronger. This provides a simple and predictable knob for users to modulate the edit's impact.

F.4. Analysis of δ and λ

We conduct a detailed ablation study on the two most critical hyperparameters of the Chord Control Field: the temporal window size δ and the step scale λ. Figure 7 presents a comprehensive quantitative analysis, visualizing the joint impact of δ and λ on semantic alignment (CLIP) and a suite of background preservation metrics (LPIPS, MSE, SSIM, and PSNR).

The 3D surface plots reveal a clear and complex trade-off that validates our core hypothesis. Across all metrics, the naive baseline ($\delta = 0$, the front edge of each plot) represents the worst-performing configuration, exhibiting the poorest semantic alignment (lowest CLIP) and the highest distortion (highest LPIPS/MSE, lowest SSIM/PSNR).

The impact of δ (temporal smoothing) is twofold:

1. **On Preservation (Monotonic):** As δ increases, we observe a robust and monotonic improvement in all background preservation metrics (LPIPS/MSE $\downarrow$, SSIM/PSNR $\uparrow$). This confirms that temporal smoothing is fundamentally key to stabilizing the field and reducing distortion.
2. **On Semantics (Non-Monotonic):** The effect of δ on semantic alignment (CLIP) is non-monotonic. Moving from $\delta = 0$ (naive) to a small δ (e.g., $\approx 0.15 - 0.3$) *improves* the CLIP score, as smoothing prevents the catastrophic semantic collapse of the naive baseline. However, as δ becomes too large (e.g., $\delta > 0.3$), the smoothing becomes "too conservative," overly suppressing the intended edit and *reducing* the CLIP score.

This explains the existence of an optimal δ "sweet spot" that balances stabilizing the edit (improving CLIP) against becoming overly conservative (harming CLIP). Concurrently, increasing λ (the step scale) serves as a more direct control for edit strength, robustly increasing semantic alignment at the expected cost of decreased background fidelity. The smoothness of these surfaces demonstrates that these parameters offer a predictable and stable trade-off.

G. More Quantitative Results

Table 1 provides the full, unabridged quantitative results that underpin the claims made in the main paper. This detailed table expands upon the main paper's Table 1 by including additional, fine-grained metrics for background and structural preservation: Structural Distance (Struct. Dist.), Structural Similarity Index (SSIM), and Perceptual Similarity (LPIPS). This data provides a comprehensive view of our method's performance and allows for a deeper analysis.

Validating the Chord Control Field. The primary claim of our work is that the naive single-step editing field ($\delta = 0$) is unstable, and our Chord Control Field ($\delta > 0$) fundamentally resolves this. The detailed metrics in Table 1 provide overwhelming evidence for this claim. When comparing our method (Ours) against the baseline (Naive) across all three models, our method consistently demonstrates superior structural and perceptual fidelity. The Naive rows show significantly higher structural distortion and perceptual error compared to our Ours rows, especially in the LPIPS and Struct. Dist. columns. This trend holds true across all preservation metrics, proving that our temporal smoothing (when $\delta > 0$) is critical for preserving background consis-

Table 1. Quantitative comparison of our method against other editing methods on PIE Bench.

Type	Method	Struct.	Background Preservation				CLIP Semantics		Efficiency			
		Dist.$_{10^3}$ ↓	PSNR↑	MSE$_{10^3}$ ↓	SSIM$_{10^2}$ ↑	LPIPS$_{10^3}$ ↓	Whole↑	Edited↑	Runtime(s)↓	Step↓	NFE↓	VRAM(MiB)↓
Multi-step (≥ 20 steps)	DDIM + MasaCtrl	28.79	21.25	8.58	80.11	106.59	24.13	21.13	55.19	50	100	12272
	Direct Inversion + MasaCtrl	24.46	21.78	7.99	81.74	87.38	24.42	21.38	79.10	50	100	12272
	DDIM + PnP	28.20	21.26	8.42	78.90	113.58	25.45	22.54	28.01	50	100	9262
	Direct Inversion + PnP	24.27	21.43	8.10	79.52	106.26	25.48	22.63	28.03	50	100	9262
	FlowEdit (SD3)	12.34	22.17	7.69	83.54	104.81	**26.64**	**23.69**	7.22	33	33	17140
Few-step (4 steps)	TurboEdit (SDXL-Turbo)	13.80	21.44	9.49	80.08	108.60	24.66	21.79	2.69	4	4	13826
	InfEdit (SD1.4)	17.06	**24.14**	6.82	**85.02**	55.69	24.89	21.88	1.41	4	4	6502
	InstantEdit (PeRFlow-SD1.5)	7.14	23.80	**4.21**	84.84	60.92	24.97	21.82	1.30	4	8	16270
One-step	SwiftEdit (SwiftBrush-v2)	12.96	21.71	8.22	74.84	91.22	24.93	21.85	0.54	**1**	2	15060
	ChordEdit (Naive, InstaFlow)	13.32	22.05	10.45	73.49	103.33	22.97	20.19	**0.38**	**1**	2	**6198**
	ChordEdit (Our, InstaFlow)	**6.33**	23.05	5.45	82.49	**53.33**	24.17	21.39	**0.38**	**1**	2	**6198**
	ChordEdit (Naive, SwiftBrush-v2)	16.33	20.52	17.17	73.42	127.43	23.78	21.06	**0.38**	**1**	2	6988
	ChordEdit (Our, SwiftBrush-v2)	12.96	22.04	7.13	75.84	111.22	25.12	22.58	**0.38**	**1**	2	6988
	ChordEdit (Naive, SD-Turbo)	25.44	21.38	9.73	74.39	131.30	25.11	21.96	**0.38**	**1**	2	6988
	ChordEdit (Naive w/o prox, SD-Turbo)	19.18	21.89	10.84	77.24	105.27	23.68	20.83	**0.20**	**1**	**1**	6988
	ChordEdit (Ours, SD-Turbo)	16.58	22.20	6.84	75.91	128.25	25.58	22.96	**0.38**	**1**	2	6988
	ChordEdit (Ours w/o prox, SD-Turbo)	10.37	23.89	5.05	81.24	88.36	24.97	21.87	**0.20**	**1**	**1**	6988

tency, not just in pixel space (PSNR), but also in structural (SSIM) and deep-feature (LPIPS) space.

Decoupling Transport (NFE=1) and Refinement (NFE=2). The full table explicitly demonstrates our framework's core design: the decoupling of consistency-preserving transport (NFE=1) from semantic-boosting refinement (+1 NFE). This addresses any concerns regarding the NFE=2 configuration. Let us compare the **Ours (w/o prox, SD-Turbo)** (NFE=1) variant against the full **Ours (SD-Turbo)** (NFE=2) variant. The data shows a clear and intentional trade-off. The **Ours (w/o prox)** variant consistently achieves the best scores across the full suite of background and structural preservation metrics, including Struct. Dist., PSNR, SSIM, and LPIPS. This is the pure, low-energy transport. By adding the optional refinement step, the full **Ours** (NFE=2) variant trades a predictable amount of this preservation to achieve a significant boost in semantic alignment, as measured by the CLIP-Edited score. This data strongly supports our claim that **ChordEdit** is a modular framework. Users can choose the NFE=1 variant for maximum fidelity and true one-step performance, or the NFE=2 variant for the best overall semantic alignment, fully vindicating the design presented in the main paper.

Efficiency and Model Agnosticism. Finally, the table confirms our claims of efficiency and broad applicability. The performance gains of **Ours** over **Naive** are consistent across **InstaFlow**, **SwiftBrush-v2**, and **SD-Turbo**, confirming the model-agnostic nature of our Chord Control Field. Furthermore, our VRAM footprint is shown to be exceptionally low, representing a significant practical advantage over memory-intensive methods like **SwiftEdit** or **FlowEdit**, making our method far more accessible for real-time applications.

H. More Qualitative Results

We provide additional qualitative comparisons in Figure 12 and Figure 13. These examples further support the quantitative findings in the main paper.

Across a diverse set of editing prompts, our method consistently produces high-fidelity results that adhere to the target prompt while maintaining exceptional background preservation. This contrasts sharply with multi-step methods, which often introduce undesirable artifacts or fail to preserve the subject's identity (e.g., *Direct Inversion+PnP*), and other few-step methods that struggle to balance semantic accuracy with structural consistency. As demonstrated, our method successfully avoids the catastrophic distortions and background collapse seen in naive one-step approaches, validating the stability of our Chord Control Field.

I. Societal Impacts

The development of high-fidelity, real-time generative models like ChordEdit presents significant opportunities while also necessitating a discussion of societal and ethical considerations.

Positive Applications and Broader Impacts. Our primary motivation for developing ChordEdit is to democratize high-end creative tools. The method's core advantages—its speed, resource efficiency (low VRAM), and model-agnostic, training-free nature—make powerful, real-time generative editing accessible to a broader audience, including those without specialized high-end hardware. We envision our work empowering artists, designers, content creators, and hobbyists by providing an intuitive and responsive tool for rapid prototyping, creative exploration, and concept visualization. For example, a designer could instantly visualize different seasonal aesthetics for a landscape (e.g., "fall" to "spring"), or

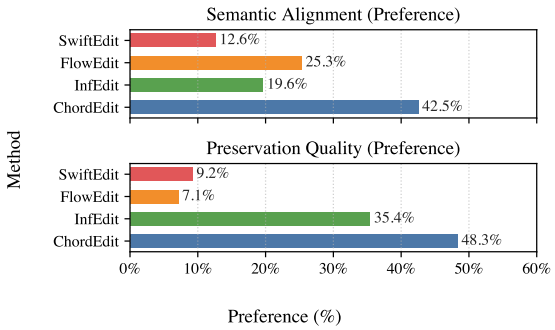

Figure 10. **User Study Results.** Aggregated human preference rates from a four-way blind comparison, matching the data cited in the main paper. ChordEdit was the clear winner in *both* Semantic Alignment (42.5%) and Preservation Quality (48.3%), demonstrating its superior overall performance.

a casual user could easily modify personal photos in real-time. This reduces the barrier to entry for complex image manipulation, fostering greater creative expression.

Ethical Considerations and Potential for Misuse. Like all high-fidelity generative models, ChordEdit carries the risk of misuse. As acknowledged in the main paper, the ability to create realistic and consistent edits in real-time could be exploited to generate deceptive content, spread misinformation, or create malicious media. The high quality and structural preservation of our method might make such forgeries more convincing.

Furthermore, ChordEdit is a training-free method that operates on pre-trained text-to-image models. It does not, by itself, correct any inherent societal biases (e.g., related to race, gender, or culture) that may be present in these foundational models. As such, edits performed by ChordEdit may reflect or even amplify these underlying biases, depending on the prompts and the backbone model used.

Mitigation and Author Statement. We strongly condemn the use of our technology for any deceptive or harmful purpose. Our work is intended for creative and assistive applications, aimed at augmenting human creativity, not replacing it or enabling deception. We believe the best mitigation strategy lies in the concurrent development of robust detection tools for synthetic media, as well as in fostering public awareness and critical media literacy. We encourage the research community to continue to prioritize the development of ethical guidelines and safeguards alongside the advancement of generative capabilities.

J. User Study

To complement our quantitative analyses, we conducted a formal user study to assess human perceptual preference.

Algorithm 1 ChordEdit (Multi-noise $n > 1$ version)

Inputs: source image x_{src}; prompts $c_{\text{src}}, c_{\text{tar}}$; step time t; window δ; step scale λ; Proximal Refinement time t_{c}; **number of noise samples** n.
Output: edited image x_{tar}.
Init: $x_{\text{in}} \leftarrow x_{\text{src}}$
$\hat{u}_{\text{sum}} \leftarrow 0$
for $i = 1$ **to** n **do**
  $\mathbf{R}_{t-\delta} \leftarrow \mathbf{R}(x_{\text{in}}, t-\delta)$
  $R_t \leftarrow R(x_{\text{in}}, t)$
  $\hat{u}_i \leftarrow \frac{t\,\mathbf{R}_{t-\delta} + \delta\,\mathbf{R}_t}{t+\delta}$
  $\hat{u}_{\text{sum}} \leftarrow \hat{u}_{\text{sum}} + \hat{u}_i$
end for
$\hat{u}_{\text{avg}} \leftarrow \hat{u}_{\text{sum}}/n$
$x^{\text{pred}} \leftarrow x_{\text{in}} + \lambda\,\hat{u}_{\text{avg}}$
$x_{\text{tar}} \leftarrow \text{prox}\left(x^{\text{pred}},\, t_{\text{c}},\, c_{\text{tar}}\right)$
Return x_{tar}

An example of the evaluation form shown to participants is provided in Figure 11. Automated metrics often fail to capture the holistic "quality" or "naturalness" of an edit, so this study was designed to validate our core claim: ChordEdit produces edits that are semantically accurate, and also more structurally consistent and artifact-free than competing methods.

For the study setup, we recruited 150 participants with diverse backgrounds. We presented them with a four-way blind comparison, showing the original image, a text prompt, and four edited results from ChordEdit, InfEdit, FlowEdit, and SwiftEdit. The order of all results was randomized to prevent bias.

We asked participants to vote for the single best image based on two independent criteria. The first was Semantic Alignment, judging which image best matched the text prompt's meaning. The second was Preservation Quality, judging which image looked most natural and best preserved the background and non-edited regions, with the fewest artifacts.

The study yielded 4,500 total votes (150 participants $\times$ 30 prompts) for each criterion. The results, shown in Figure 10, confirm the data cited in the main paper and show a clear preference for ChordEdit.

For Semantic Alignment, ChordEdit was the clear winner, preferred in **42.5%** of comparisons. This significantly outperformed FlowEdit (25.3%), while InfEdit (19.6%) and SwiftEdit (12.6%) lagged considerably.Proposition 4

For Preservation Quality, ChordEdit also achieved the top position, securing **48.3%** of the vote. This result is particularly compelling, as it shows our method was perceived as more stable and artifact-free than even InfEdit (35.4%), a baseline renowned for its high preservation. FlowEdit (7.1%) and SwiftEdit (9.2%) were frequently penalized by

users for artifacts and distortion.

In conclusion, the user study strongly validates our claims. ChordEdit was the only method to rank first in both categories. Other methods force a compromise—excelling at either semantics (FlowEdit) or preservation (InfEdit) but failing at the other. ChordEdit was the most preferred for both, confirming it provides the best overall perceptual quality and proving our low-energy, stable transport field translates directly to the most desirable result for human observers.

K. ChordEdit algorithm with multi-noise

For maximum efficiency, our core ChordEdit algorithm presented in the main paper uses a single noise sample ($n = 1$). However, our framework can be directly extended to support multiple Monte Carlo (MC) noise samples ($n > 1$) to theoretically further reduce the estimation variance.

Algorithm 1 provides the pseudocode for this multi-noise version. The key difference is that we first compute the Chord Control Field $\hat{u}_i$ independently for n different noise samples, then average these fields, and finally use this averaged field $\hat{u}_{\text{avg}}$ to perform the single-step transport and subsequent proximal refinement.

L. Symbols Table

Table 2. Symbols Table

Symbol	Description
$t \in [0, 1]$	Time variable for the probability flow. $t = 1$ corresponds to the source data distribution, and $t = 0$ to the target noise distribution.
x_t	The image state at time t.
c	The text condition (prompt).
$c_{\text{src}}, c_{\text{tar}}$	The source and target text prompts, respectively.
$x_{\text{src}}, x_{\text{tar}}$	The source image (x_1) and the desired target image (x_0), respectively.
$p_t(x \mid c)$	The probability distribution of x_t at time t conditioned on c.
p_1, p_0	The source data distribution ($t = 1$) and the target noise distribution ($t = 0$).
z	A synthetic noisy proxy state used to query the model.
x_τ	The editing anchor, fixed to the clean source image x_1 (i.e., $x_\tau := x_{\text{src}}$).
$v(x_t, t, c)$	The drift of the conditional probability flow induced by the pre-trained T2I model.
$\Delta v(x_t, t)$	The instantaneous residual field, defined as $v(\cdot, c_{\text{tar}}) - v(\cdot, c_{\text{src}})$.
$Q(z, t, c)$	The model's observable output (e.g., noise prediction, velocity) at noisy state z.
$\Delta Q(z, t)$	The conditional residual of the observable, $Q(\cdot, c_{\text{tar}}) - Q(\cdot, c_{\text{src}})$.
$\hat{\epsilon}_\theta, \mathbf{v}_\theta$	The predicted noise and predicted velocity, respectively, from a model.
$K_t(\cdot \mid x_\tau)$	The forward noising kernel that maps the anchor x_τ to a noisy state z at time t.
$\mathcal{B}_t$	A time-only linear map that projects the model's output Q into a unified comparison domain (velocity units).
$A_t^{(\epsilon)}, A_t^{(x_0)}, A_t^{(v)}, A_t^{(\text{score})}$	Specific coefficients defining $\mathcal{B}_t$ for noise, x_0, v-, and score-to-drift parameterizations.
$\alpha(t), \sigma(t)$	Coefficients of the noise schedule for the forward path $x_t = \alpha(t)x_0 + \sigma(t)\epsilon$.
$\beta(t)$	Continuous-time noise schedule parameter (related to $\alpha(t)$).
$u_t(x)$	The ideal, low-energy editing vector field that drives the transport from ρ_1 to ρ_0.
$\mathbf{R}(x_\tau, t)$	The observable proxy field; the expected value of the mapped observable residual ($\mathbb{E}[\mathcal{B}_t \Delta Q]$).
$\varepsilon_t, \eta(t)$	Zero-mean noise terms in the measurement model $\mathbf{R} = u_t + \varepsilon_t$.
u_{nai}	The naive control field, which simply uses the proxy field: $u_{\text{nai}} = \mathbf{R}$.
$\hat{u}_t(x_\tau)$	The Chord Control Field: a locally smoothed, low-energy estimator for u_t.
δ	The temporal window size for smoothing, $[t - \delta, t]$.
$\Phi_t(u; x_\tau)$	The strictly convex quadratic surrogate objective minimized to find the Chord Control Field.
$u_t^\star(x_\tau)$	The exact, integral-form minimizer of Φ_t.
$K_\delta(s)$	The causal smoothing kernel that defines $\hat{u}$ as a convolution: $\hat{u} \approx K_\delta * \mathbf{R}$.
λ	The step scale, controlling the magnitude of the applied edit transport.
x^{pred}	The predicted image after the single transport step.
$\text{prox}(\cdot)$	The optional proximal refinement step.
t_c	The time parameter used for the proximal refinement step.
n	The number of Monte Carlo noise samples used in the estimation.
$\rho_t(x)$	The transport density, evolving from $\rho_1 = p_1(\cdot \mid c_{\text{src}})$ to $\rho_0 = p_0(\cdot \mid c_{\text{tar}})$.
$\mathcal{E}[u; \rho]$	The Benamou–Brenier kinetic energy functional, $\int \int \frac{1}{2} \|u_t(x)\|^2 \rho_t(x) dx dt$.
$\bar{E}$	The discrete, unweighted Benamou–Brenier kinetic energy.
$u^\star$	The true, energy-minimizing optimal control field.
h	The step size for numerical integration (in one-step editing, $h = 1$).
τ_{n+1}	The one-step local truncation error of the numerical integrator.
$f(x, t)$	The editing vector field of the ODE, $f(x, t) = u(x, t)$.
M_f	Bound on the derivatives of f, related to local error.
$\mathcal{C}(u; \mathcal{U})$	A computable proxy for the consistency constant.

Continued on next page

Table 2 – continued from previous page

Symbol	Description
$C_{\text{cho}}, C_{\text{nai}}$	The consistency constants of the underlying ODE for the Chord and Naive fields.
L_u, M_u	Bounds on the spatial and temporal derivatives of u, used for global error analysis.
e_n^u	The global error of the numerical solution at step n.
P_δ	The L^2_ρ-orthogonal projection onto the subspace of "chord" functions.
S	The total number of discrete integration steps (step count).
s	The step index, $s \in \{1, \ldots, S\}$.

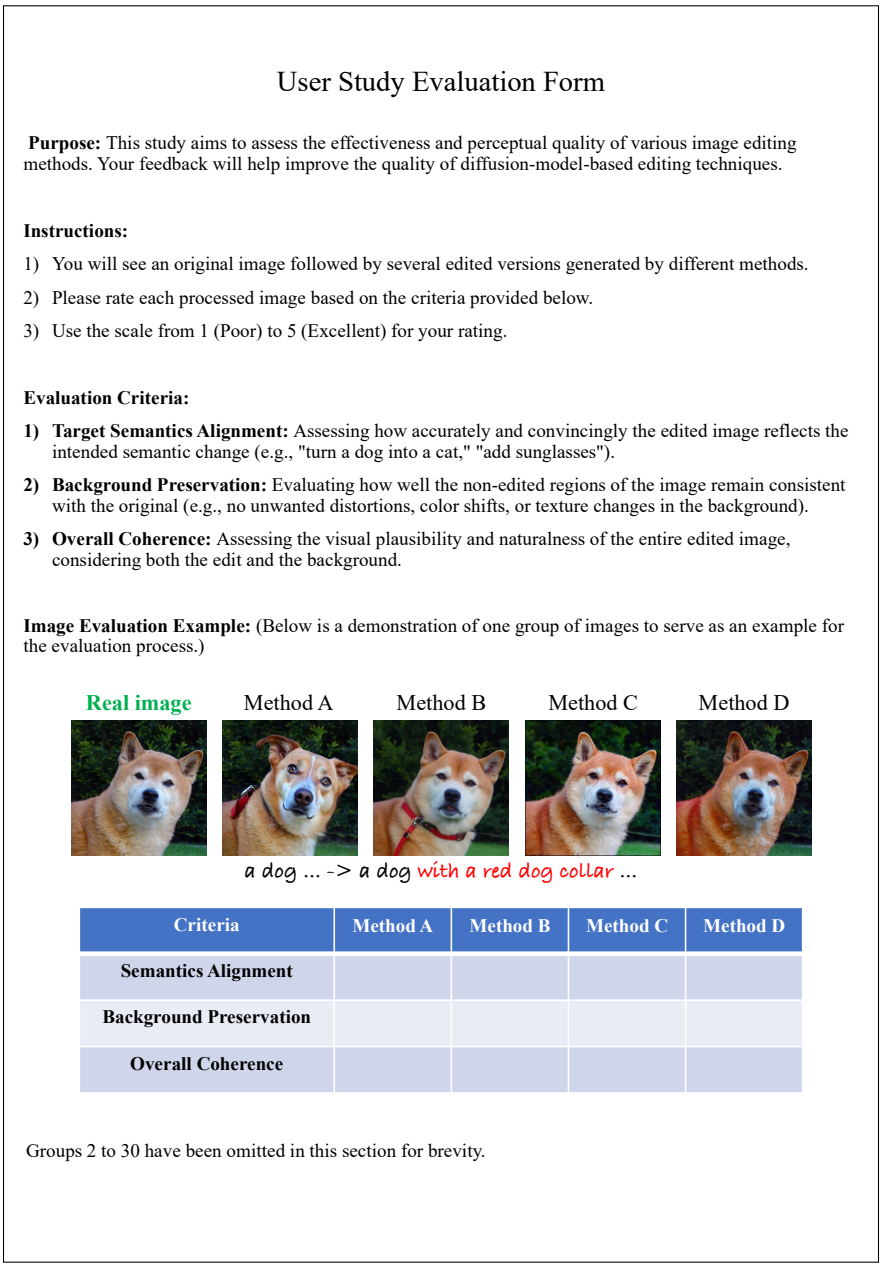

Figure 11. **User Study Form Example.**

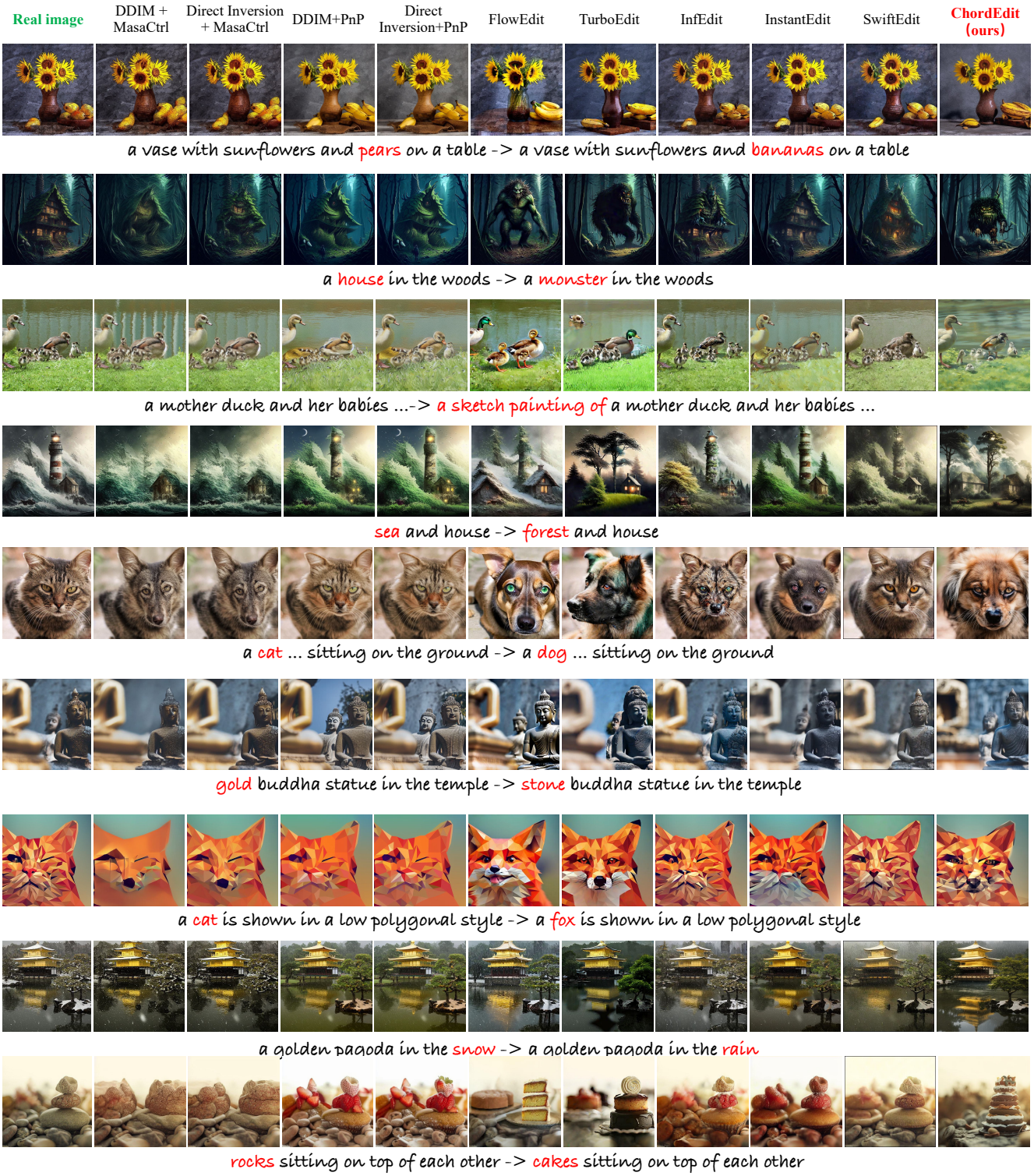

Figure 12. Comparison of Methods.

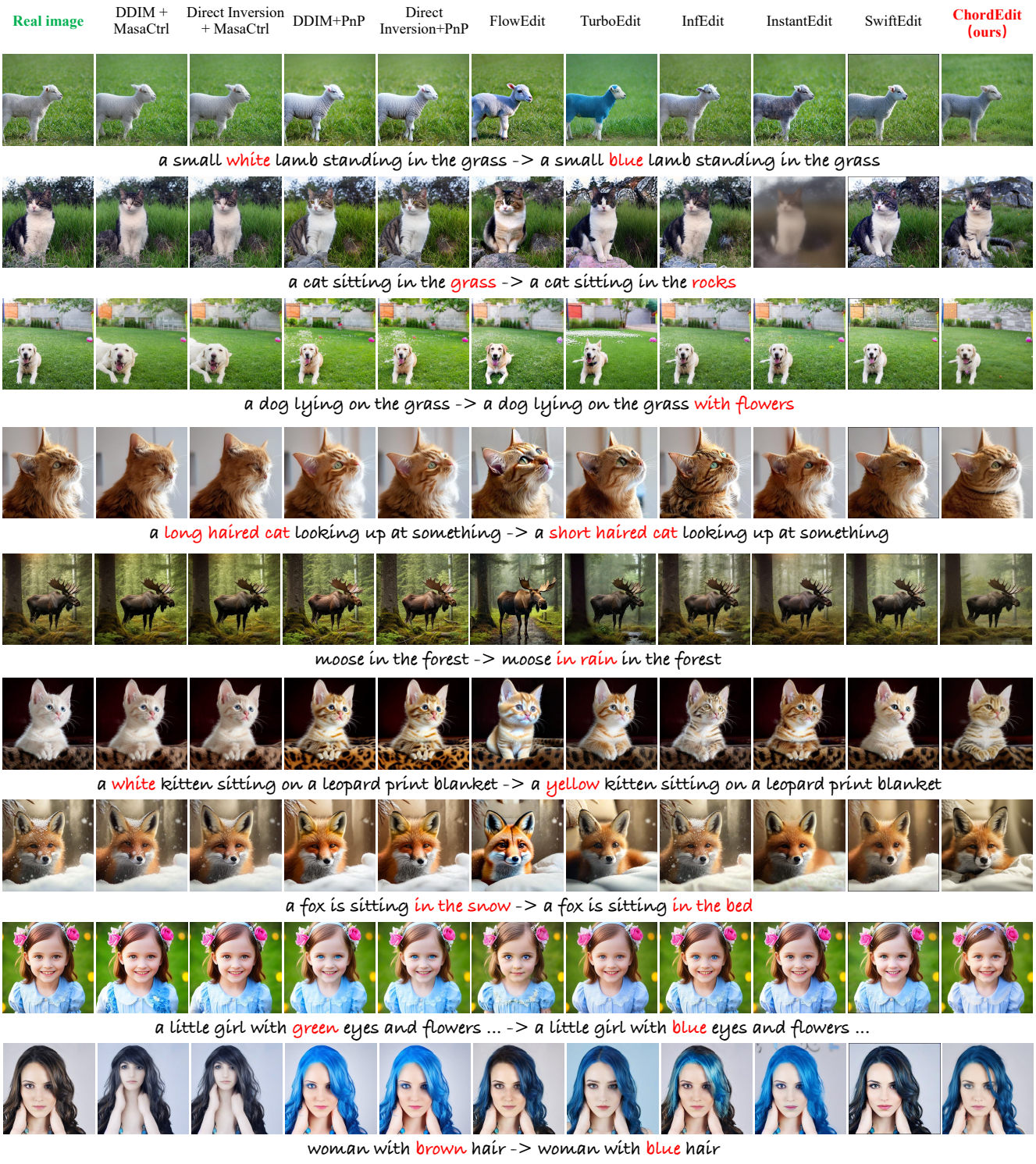

Figure 13. Comparison of Methods.

References

[1] Amir Alimohammadi, Aryan Mikaeili, Sauradip Nag, Negar Hassanpour, Andrea Tagliasacchi, and Ali Mahdavi Amiri. Cora: Correspondence-aware image editing using few-step diffusion. In *SIGGRAPH*, 2025. 1

[2] Guillaume Couairon, Jakob Verbeek, Holger Schwenk, and Matthieu Cord. Diffedit: Diffusion-based semantic image editing with mask guidance. *arXiv:2210.11427*, 2022. 1

[3] Trung Dao, Thuan Hoang Nguyen, Thanh Le, Duc Vu, Khoi Nguyen, Cuong Pham, and Anh Tran. Swiftbrush v2: Make your one-step diffusion model better than its teacher. In *Computer Vision – ECCV 2024*, pages 176–192. Springer, 2025. 1

[4] Gilad Deutch, Rinon Gal, Daniel Garibi, Or Patashnik, and Daniel Cohen-Or. Turboedit: Text-based image editing using few-step diffusion models. In *SIGGRAPH Asia 2024 Conference Papers*, pages 1–12, 2024. 1

[5] Yiming Gong, Zhen Zhu, and Minjia Zhang. Instantedit: Text-guided few-step image editing with piecewise rectified flow. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16808–16817, 2025. 1

[6] Bo Han, Yujun Shen, Yingqing He, Lei Zhang, and Jingren Zhou. Proxedit: Improving tuning-free real image editing with proximal guidance. In *WACV*, 2024. 1

[7] Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Prompt-to-prompt image editing with cross attention control. *arXiv preprint arXiv:2208.01626*, 2022. 1

[8] Xuan Ju, Ailing Zeng, Yuxuan Bian, Shaoteng Liu, and Qiang Xu. Pnp inversion: Boosting diffusion-based editing with 3 lines of code. *International Conference on Learning Representations (ICLR)*, 2024. 1

[9] Bahjat Kawar, Shiran Zada, Oran Lang, Omer Tov, Huiwen Chang, Tali Dekel, Inbar Mosseri, and Michal Irani. Imagic: Text-based real image editing with diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 1

[10] Vladimir Kulikov, Matan Kleiner, Inbar Huberman-Spiegelglas, and Tomer Michaeli. Flowedit: Inversion-free text-based editing using pre-trained flow models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19721–19730, 2025. 1

[11] Ruibin Li, Yujun Shen, et al. Source prompt disentangled inversion for boosting image editability with diffusion models. In *ECCV*, 2024. 1

[12] Shanchuan Lin, Anran Wang, and Xiao Yang. Sdxl-lightning: Progressive adversarial diffusion distillation. *arXiv preprint arXiv:2402.13929*, 2024. 1

[13] Xingchao Liu, Xiwen Zhang, Jianzhu Ma, Jian Peng, and Qiang Liu. Instaflow: One step is enough for high-quality diffusion-based text-to-image generation. In *International Conference on Learning Representations*, 2024. 1

[14] 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. *arXiv preprint arXiv:2108.01073*, 2021. 1

[15] Daiki Miyake, Akihiro Iohara, Yu Saito, and Toshiyuki Tanaka. Negative-prompt inversion: Fast image inversion for editing with text-guided diffusion models. In *WACV*, 2025. 1

[16] Ron Mokady, Amir Hertz, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Null-text inversion for editing real images using guided diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6038–6047, 2023. 1

[17] Trong-Tung Nguyen, Quang Nguyen, Khoi Nguyen, Anh Tran, and Cuong Pham. Swiftedit: Lightning fast text-guided image editing via one-step diffusion. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 21492–21501, 2025. 1

[18] Zhizhong Pan, Yijun Li, Xue Bai, Zhuowen Tu, and Ming-Hsuan Yang. Effective real image editing with accelerated iterative diffusion inversion. In *ICCV*, 2023. 1

[19] Axel Sauer, Dominik Lorenz, Andreas Blattmann, and Robin Rombach. Adversarial diffusion distillation. In *Computer Vision – ECCV 2024*, pages 87–103. Springer, 2024. 1

[20] Qi Si, Bo Wang, and Zhao Zhang. Contrastive learning guided latent diffusion model for image-to-image translation. *arXiv preprint arXiv:2503.20484*, 2025. 1

[21] Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *International Conference on Machine Learning*, pages 32211–32252. PMLR, 2023. 1, 4

[22] Nikita Starodubcev, Mikhail Khoroshikh, Artem Babenko, and Dmitry Baranchuk. Invertible consistency distillation for text-guided image editing in around 7 steps. *arXiv:2406.14539*, 2024. 1

[23] Narek Tumanyan, Michal Geyer, Shai Bagon, and Tali Dekel. Plug-and-play diffusion features for text-driven image-to-image translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1921–1930, 2023. 1

[24] Bram Wallace, Akash Gokul, and Nikhil Naik. Edict: Exact diffusion inversion via coupled transformations. In *CVPR*, 2023. 1

[25] Pengcheng Xu, Qingnan Fan, Fei Kou, Shuai Qin, Hong Gu, Ruoyu Zhao, Charles Ling, and Boyu Wang. Textualize visual prompt for image editing via diffusion bridge. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025. 1

[26] Sihan Xu, Yidong Huang, Jiayi Pan, Ziqiao Ma, and Joyce Chai. Inversion-free image editing with natural language. *CoRR*, 2023. 1

[27] Xinrui Zu and Qian Tao. Cot flow: Learning optimal-transport image sampling and editing by contrastive pairs. *arXiv preprint arXiv:2406.12140*, 2024. 2