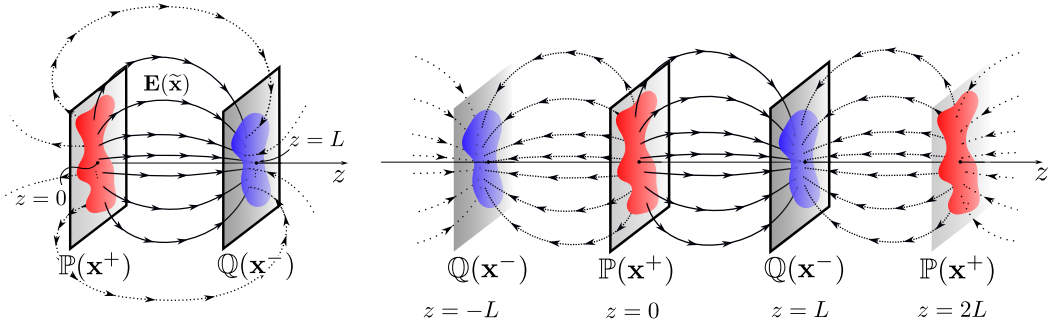
CAPACITOR CHAIN FIELD MATCHING FOR ELECTROSTATIC GENERATIVE MODELING

Anonymous authors
Paper under double-blind review

ABSTRACT

Electrostatic Field Matching (EFM) constructs data-to-data generative transport by placing source and target distributions on oppositely charged plates in an augmented space. However, its single-capacitor geometry produces backward-oriented and exterior field lines, requiring stochastic branching for exact transport and complicating practical ODE-based inference. We propose Capacitor Chain Field Matching (C^{2}FM), a framework that addresses these geometric limitations and revisits how electrostatic transport fields are learned. First, C^{2}FM replaces a single pair of charged plates with an infinite periodic chain of alternating source and target distributions. The resulting symmetry confines field lines between adjacent plates. We prove that the ideal construction transports the source distribution to the target along forward-oriented field lines. Second, we derive a training objective with single-sample targets and theoretically grounded perturbation kernels, replacing EFM's multi-sample field estimates and heuristic perturbation scheme. Together, these contributions preserve EFM's electrostatic interpretation while enabling practical transport and training without multi-sample field estimation.



(a) EFM Kolesov et al. (2025)

(b) Our C^{2}FM

Figure 1: Electrostatic Field Matching (Kolesov et al., 2025, EFM) and our Capacitor Chain Field Matching (C^{2}FM) framework. **(a)** In EFM, two D-dimensional data distributions $\mathbb{P}(\cdot)$ and $\mathbb{Q}(\cdot)$ are embedded in $\mathbb{R}^{D+1}$ at $z = 0$ and $z = L$, respectively. The first distribution is assigned a positive charge, the second a negative charge, creating a capacitor-like electric field. Movement along the resulting field lines transports one distribution into the another one. Nevertheless, in EFM some field lines leave the region $z \in [0, L]$: both backward- and forward-oriented lines arise, and a correct transport must account for both families. As a consequence, EFM faces difficulties in training and inference (see §2.2 for details). **(b)** In C^{2}FM, we introduce infinite number of the positive and negative distribution copies. $\mathbb{P}(\cdot)$ is placed on the planes $z = 0, \pm 2L, \pm 4L, \ldots$, while $\mathbb{Q}(\cdot)$ is placed on $z = \pm L, \pm 3L, \pm 5L, \ldots$, forming a periodic chain of capacitors. This charge configuration is symmetric under reflection across every plane $z = nL,\ n \in \mathbb{Z}$, which resolves the principal limitations of EFM while staying fully within the electrostatic paradigm (see §3.1).

1 INTRODUCTION

Diffusion models (Song et al., 2020; Karras et al., 2022), Flow Matching (Lipman et al., 2023; Liu et al., 2022), and Stochastic Interpolants (Albergo & Vanden-Eijnden, 2023; Albergo et al., 2025) have become standard tools for both noise-to-data generation and data-to-data transport. These methods typically learn a velocity field whose integral curves transform one distribution into another.

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.

A complementary line of work proposes to construct similar generative fields from electrostatic principles (Xu et al., 2023; Shlenskii & Korotin, 2026). Poisson Flow Generative Models (Xu et al., 2022, PFGM) interpret the data distribution as an electric charge distribution embedded in an augmented space. The corresponding electrostatic field, obtained by the superposition principle, defines field lines that connect a simple prior distribution to the data distribution. This provides a physics-inspired mechanism for noise-to-data generation, in which samples are generated by following electrostatic field lines.

Electrostatic Field Matching (Kolesov et al., 2025, EFM) extends this idea to data-to-data transport. Given two distributions, EFM places them on two parallel hyperplanes in an augmented space and assigns them opposite electric charges. The resulting electrostatic field induces field lines between the two charged plates, thereby defining a transport from the source to the target distribution. This construction is theoretically appealing because the electrostatic field is determined solely by the source and target marginals through superposition, rather than by a prescribed coupling between samples.

However, the single-capacitor geometry of EFM introduces several practical difficulties (Figure 1a). Field lines are not confined to the slab between the plates: backward-oriented and exterior trajectories arise, making the exact transport stochastic and complicating ODE-based inference. In addition, practical EFM regresses the normalized field direction from finite-sample Monte Carlo estimates at points drawn from an auxiliary training-volume distribution. Thus, its learning procedure depends both on the choice of this distribution and on the quality of the finite-sample field estimate.

Interaction Field Matching (Manukhov et al., 2026, IFM) was proposed as a general field-based framework that contains EFM as a special case. Within this framework, one can choose interaction fields with more convenient field-line geometry, thereby avoiding some of the practical difficulties of EFM. However, recent work (Shlenskii et al., 2026) has shown that certain practical IFM constructions coincide with existing flow-based stochastic interpolants, and thus no longer preserve the electrostatic structure of EFM. This raises a question: can one overcome the practical limitations of EFM while preserving its electrostatic nature?

In this work, we answer this question affirmatively. We propose $\mathrm{C}^2\mathrm{FM}$ (Figure 1b), which replaces the single capacitor of EFM by a periodic chain of alternating source and target copies located at $z = 2nL$ and $z = (2n+1)L$, respectively. The resulting reflection symmetry confines forward field lines between adjacent plates, yielding a deterministic transport while preserving the electrostatic interpretation of EFM.

Moreover, we represent the finite-chain field direction through four Coulomb-weighted conditional moments and a single mixing logit, and recover the required moments using Coulomb-kernel sampling and denoising objectives rather than direct regression of a finite-sample normalized field.

Our **contributions** are as follows:

- We introduce $\mathrm{C}^2\mathrm{FM}$, a periodic electrostatic construction whose field is forward-oriented inside the slab and whose deterministic field-line map transports the source distribution to the target distribution (§3.1).
- We prove uniform convergence of the finite capacitor chain to the ideal infinite-chain field and give conditions under which the field approximation controls the induced transport (§3.2).
- We derive an exact representation of the finite-chain field direction through Coulomb-weighted conditional moments and develop a corresponding population-consistent denoising procedure (§5.1).

2 Background

This section reviews the electrostatic concepts used throughout the paper. We first recall the superposition principle and field lines in §2.1, then summarize EFM and its limitations in §2.2.

2.1 Electrostatics

We start with the fundamental principles of electrostatics necessary for understanding electrostatic-based generative models. A detailed treatment of three-dimensional electrostatics can be found in any standard electricity and magnetism textbook, e.g., (Landau & Lifshitz, 1971, Chapter 5). The generalization of electrostatics to high-dimensional spaces is discussed in (Caruso et al., 2023) and (Kolesov et al., 2025, Appendix A).

The electrostatic field. Let $q : \mathbb{R}^D \to \mathbb{R}$ be the density of a charge distribution on $\mathbb{R}^D$. The distribution may contain both positive and negative charges and is assumed to have finite total charge ($\int |q(\mathbf{x})| d\mathbf{x} < \infty$). At a point $\mathbf{x} \in \mathbb{R}^D$ it produces the electrostatic field $\mathbf{E} : \mathbb{R}^D \to \mathbb{R}^D$:

$$\mathbf{E}(\mathbf{x}) = \int \frac{1}{S_{D-1}} \frac{(\mathbf{x} - \mathbf{x}')}{||\mathbf{x} - \mathbf{x}'||^D} q(\mathbf{x}') d\mathbf{x}', \tag{1}$$

where S_{D-1} is the surface area of an $(D-1)$-dimensional sphere with unit radius. Equation (1) represents the field as a superposition of point-charge fields.

Electric field lines. An electric field strength line is a curve $\mathbf{x}(\tau) \in \mathbb{R}^D,\ \tau \in [a, b] \subset \mathbb{R}$ whose tangent to each point is parallel to the electric field at that point. In other words:

$$\frac{d\mathbf{x}(\tau)}{d\tau} \sim \mathbf{E}(\mathbf{x}). \tag{2}$$

Electric field lines are a key concept for electrostatic generative models such as PFGM and EFM.

2.2 Electrostatic Field Matching (EFM)

The first generative models based on electrostatics, Poisson Flow Generative Models (PFGM and PFGM++) (Xu et al., 2022; 2023), addressed the problem of mapping a simple noise distribution to a complex data distribution. To handle data-to-data transport, Electrostatic Field Matching (EFM) (Kolesov et al., 2025) extends the idea by introducing charge distributions of opposite signs. In EFM, two data distributions $\mathbb{P}(\mathbf{x}^+)$ and $\mathbb{Q}(\mathbf{x}^-)$ on $\mathbb{R}^D$ are embedded into an extended space $\mathbb{R}^{D+1}$ as charged hyperplanes: positive copies $\mathbb{P}$ are placed on the plane $z = 0$ and negative copies $\mathbb{Q}$ on $z = L$. This arrangement mimics a $(D+1)$-dimensional **capacitor**.

Points in the extended space are written as $\widetilde{\mathbf{x}} = (\mathbf{x}, z) \in \mathbb{R}^{D+1}$. The total electric field is the superposition of the fields generated by the two plates:

$$\mathbf{E}(\widetilde{\mathbf{x}}) = \mathbf{E}^+(\widetilde{\mathbf{x}}) + \mathbf{E}^-(\widetilde{\mathbf{x}}), \tag{3}$$

where $\mathbf{E}^\pm$ are obtained from the charge densities $\mathbb{P}$ and $\mathbb{Q}$ via the high-dimensional Coulomb integral (1) in $D+1$ dimensions.

The field lines generated by Eq. (3) transport $\mathbb{P}$ to $\mathbb{Q}$ almost surely (Kolesov et al., 2025). In practice, EFM learns their direction by replacing the Coulomb integrals with finite-sample Monte Carlo estimates. Denoting the resulting field estimate by $\widehat{\mathbf{E}}_M(\widetilde{\mathbf{x}})$, the regression objective is

$$\mathbb{E}_{\widetilde{\mathbf{x}} \sim p_{\text{vol}}} \left\| f_\theta(\widetilde{\mathbf{x}}) - \frac{\widehat{\mathbf{E}}_M(\widetilde{\mathbf{x}})}{\|\widehat{\mathbf{E}}_M(\widetilde{\mathbf{x}})\|} \right\|^2 \to \min_\theta, \tag{4}$$

where p_{vol} samples the training volume in the augmented space.

Limitations of EFM. Despite providing a valid transport, EFM has several practical limitations:

1. **Field lines leaving $[0, L]$.** The electric field generated by the two charged plates contains both forward-oriented trajectories, which initially move from the source plate toward increasing z, and backward-oriented trajectories, which initially move in the opposite direction. Moreover, some forward trajectories cross the target plane $z = L$ before eventually returning to it (Figs. 1a, 2).

2. **Stochasticity of the transport map.** Both forward- and backward-oriented families are required to fully recover the target distribution, so the exact EFM transport involves stochastic choices between field-line branches (Kolesov et al., 2025, §3.2).

3. **Inference difficulties.** Practical EFM uses only forward trajectories (Kolesov et al., 2025, Algorithms 1, 2); their possible overshoot beyond $z = L$ complicates ODE integration and requires heuristic stopping.

4. **Difficulties of direct field regression.** The sampling distribution p_{vol} is chosen heuristically and must cover the regions visited by the trajectories. Moreover, the target direction is obtained by normalizing a finite-sample Monte Carlo field estimate. Since normalization is nonlinear, even an unbiased $\widehat{\mathbf{E}}_M$ does not in general yield an unbiased estimate of $\mathbf{E}/\|\mathbf{E}\|$. Related multi-sample field estimators can also concentrate on only a few samples in high dimensions (Shlenskii et al., 2026).

3 CAPACITOR CHAIN FIELD MATCHING (C^2FM)

To address the limitations of EFM, we propose a new charge arrangement geometry. Unlike EFM, where the distributions $\mathbb{P}(\widetilde{\mathbf{x}})$ and $\mathbb{Q}(\widetilde{\mathbf{x}})$ are placed on the planes $z = 0$ and $z = L$ respectively, we propose creating an infinite number of copies of $\mathbb{P}(\widetilde{\mathbf{x}})$ and $\mathbb{Q}(\widetilde{\mathbf{x}})$. The positive charge distributions are placed on the planes $z = 0, \pm 2L, \pm 4L, \dots$ while the negative ones on $z = \pm L, \pm 3L, \pm 5L, \dots$ (see Fig. 1). The resulting periodic configuration is reflection symmetric with respect to every plane $z = nL$. As shown below, its axial field is strictly positive in the slab $0 < z < L$, which confines the forward transport between adjacent source and target plates.

3.1 MAIN THEORETICAL CONSTRUCTION

Let $\mathbb{P}(\mathbf{x})$ and $\mathbb{Q}(\mathbf{x})$, with $\mathbf{x} \in \mathbb{R}^D$, be two data distributions. We assign to the first distribution a positive charge and to the second distribution a negative charge. Their periodic arrangement in $\mathbb{R}^{D+1}$ can be written as

$$\begin{aligned} q^+(\widetilde{\mathbf{x}}) = q^+(\mathbf{x}, z) &= \mathbb{P}(\mathbf{x}) \sum_{n=-\infty}^{\infty} \delta(z - 2nL), \\ q^-(\widetilde{\mathbf{x}}) = q^-(\mathbf{x}, z) &= -\mathbb{Q}(\mathbf{x}) \sum_{n=-\infty}^{\infty} \delta(z - (2n+1)L). \end{aligned} \tag{5}$$

Since this infinite array has infinite total charge, we define its field through symmetric finite-chain truncations. Let $\mathbf{E}_n^+$ and $\mathbf{E}_n^-$ denote the fields generated by the n-th positive and negative plates:

$$\begin{aligned} \mathbf{E}_n^+(\widetilde{\mathbf{x}}) &= \int \frac{\big(\mathbf{x} - \mathbf{x}',\, z - 2nL\big)\mathbb{P}(\mathbf{x}')\, d^D\mathbf{x}'}{S_D\big(\|\mathbf{x} - \mathbf{x}'\|^2 + (z - 2nL)^2\big)^{\frac{D+1}{2}}}, \\ \mathbf{E}_n^-(\widetilde{\mathbf{x}}) &= \int \frac{-\big(\mathbf{x} - \mathbf{x}',\, z - (2n+1)L\big)\mathbb{Q}(\mathbf{x}')\, d^D\mathbf{x}'}{S_D\big(\|\mathbf{x} - \mathbf{x}'\|^2 + (z - (2n+1)L)^2\big)^{\frac{D+1}{2}}}, \end{aligned} \tag{6}$$

where S_D is the surface area of a D-dimensional unit sphere. We then set

$$\mathbf{E}_C(\widetilde{\mathbf{x}}) = \sum_{n=-C}^{C} \left(\mathbf{E}_n^+(\widetilde{\mathbf{x}}) + \mathbf{E}_n^-(\widetilde{\mathbf{x}})\right), \tag{7}$$

$$\mathbf{E}(\widetilde{\mathbf{x}}) = \lim_{C \to \infty} \mathbf{E}_C(\widetilde{\mathbf{x}}). \tag{8}$$

The existence and convergence of this limit for compactly supported $\mathbb{P}$ and $\mathbb{Q}$ are established in §3.2.

Theorem 3.1 (Only forward-oriented field lines)**.** *Let $\mathbb{P}$ and $\mathbb{Q}$ have compact support and let $\mathbf{E}$ be the infinite-chain field defined in Eq.* (8)*. Then*

$$E_z(\mathbf{x}, z) > 0, \qquad \mathbf{x} \in \mathbb{R}^D, \quad 0 < z < L.$$

Consequently, every forward field line starting from the source plate $z = 0$ remains in the slab $0 \le z \le L$ and is strictly forward-oriented in its interior.

The proof is given in Appendix A. For $\mathbb{P}$-almost every $\mathbf{x}^+$, consider the forward field line starting at $(\mathbf{x}^+, 0)$. By Theorem 3.1 and the standard properties of electrostatic field lines (Kolesov et al., 2025, Appendix A), it terminates on the negative plate at a point $(\mathbf{x}^-, L)$. We define the deterministic field-line map by $T(\mathbf{x}^+) = \mathbf{x}^-$.

Theorem 3.2 (**C^2FM transport theorem**)**.** *Let $\mathbb{P}$ and $\mathbb{Q}$ be compactly supported distributions and let T be the field-line map defined above. Then $T_{\#}\mathbb{P} = \mathbb{Q}$.*

Thus, unlike EFM, C^2FM defines a deterministic transport between the two distributions. The proof of Theorem 3.2 is given in Appendix A.

3.2 FINITE CAPACITOR APPROXIMATION

The infinite-chain field introduced above is defined as the limit of the symmetrically truncated fields $\mathbf{E}_C$. Since the contribution of distant capacitor cells decays with their axial distance, the truncated field $\mathbf{E}_C$ converges uniformly to the infinite-chain field $\mathbf{E}$. In particular, for compactly supported $\mathbb{P}$ and $\mathbb{Q}$ we obtain

$$\sup_{\mathbf{x} \in \mathbb{R}^D,\, 0 < z < L} \|\mathbf{E}_C(\mathbf{x}, z) - \mathbf{E}(\mathbf{x}, z)\| = O(C^{-D}). \tag{9}$$

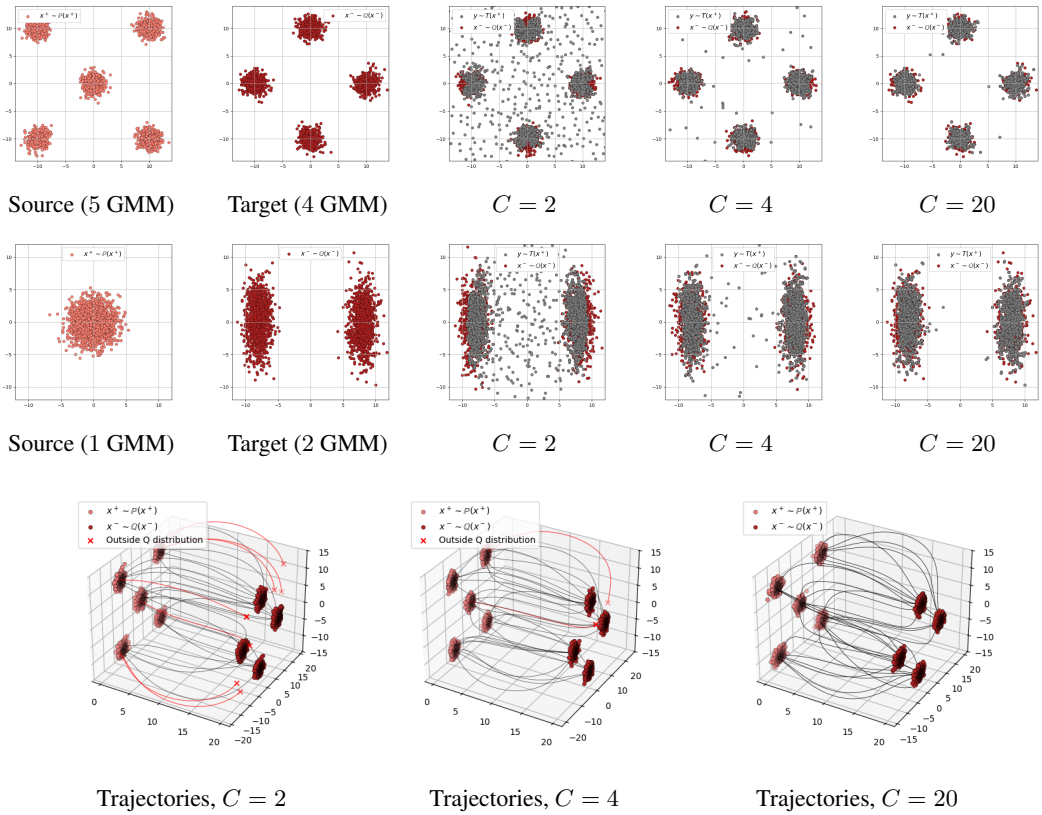


Figure 2: Qualitative results on two-dimensional Gaussian mixture transport tasks. First row: transport from a 5-component GMM to a 4-component GMM. Second row: transport from a single Gaussian to a 2-component GMM. Third row: transport trajectories for the $5 \to 4$ task. Increasing the number of capacitors C improves sample quality and produces more stable trajectories. Trajectories that extend beyond the two planes region are highlighted in red.

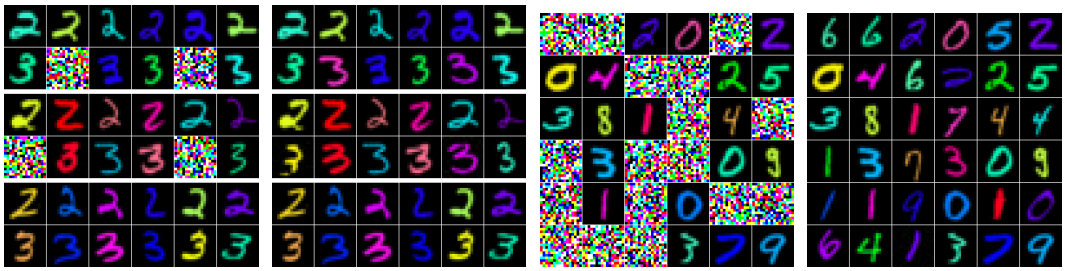
A small field error alone does not automatically imply a small error of the induced transport. However, if the ideal trajectory has a strictly positive forward-field component and the corresponding field-line dynamics is locally Lipschitz in a neighborhood of the trajectory, the finite-chain trajectory remains close to the infinite-chain one. Explicit bounds on both the field truncation error and the resulting endpoint error, together with their proofs, are given in Appendices B and C.

3.3 Validating the Construction

Before any network is introduced we test the construction on its own. The only expectations in $\mathbf{E}_C$ are over $\mathbb{P}$ and $\mathbb{Q}$; replacing them by sums over samples makes the superposition finite, and the field can then be evaluated and integrated directly. Nothing is fitted, so whatever we observe belongs to the construction.

Figure 2 sweeps C on two Gaussian-mixture tasks, one that merges and rearranges modes ($5 \to 4$) and one that splits them ($1 \to 2$). As C grows, the generated points align with the target and the trajectories straighten, as the rate (9) predicts. Lines that miss the target leave through the target plate and are highlighted in red in the third row: wide and strongly curved. For EFM the exterior field carries such a line back to the target plate from the far side, which is the backward-oriented family EFM has to account for (§2.2).

Figure 3 repeats the comparison at $D = 768$ on colored MNIST, for the translation $2 \to 3$ and for generation from noise, contrasting $C = 0$ with the chain at $C = 32$, which at this exponent is the untruncated chain to double precision (§F.2). A missed line is easy to recognise here: it ends far outside the data, and the cell it produces is saturated noise rather than a digit. Such cells fill much of the $C = 0$ grid on generation and a smaller part of it on translation, and with the chain they are gone from both.



(a) $C = 0$ (b) $C = 32$ (c) $C = 0$ (d) $C = 32$

Figure 3: Qualitative results of the exact C^{2}FM field on colored MNIST $2 \to 3$ translation **(a, b)** and noise $\to$ colored MNIST generation **(c, d)**, for a single capacitor ($C = 0$) and for the chain ($C = 32$). The field is assembled directly from the empirical measures, with no network in the loop, so these panels isolate the construction from the learning. In **(a, b)** rows alternate a source digit and its image. At $C = 0$ a line that leaves the slab ends far outside the data and renders as noise; with the chain such cells are gone.

Two things follow, and both concern the construction alone. The chain is what makes the field transport: Figure 3 sets it against the single capacitor of EFM at $C = 0$, and Figure 2 shows the approach to the limit as C grows, with the estimator, the integrator and the data fixed throughout. And the effect is not a low-dimensional curiosity, since it holds at $D = 768$ in both regimes.

4 Relation to Other Methods

C^{2}FM is an electrostatic generative model. The first works in this direction, (Xu et al., 2022, PFGM) and (Xu et al., 2023, PFGM++), generalize diffusion models, but like early diffusion models they are limited to noise-to-data generation. To handle data-to-data transfer, (Kolesov et al., 2025, EFM) was introduced; the method itself and its limitations are discussed in §2.2.

A subsequent approach, (Manukhov et al., 2026, Interaction Field Matching (IFM)), addresses the problems of EFM by broadening the electrostatic perspective to the general framework of *interaction fields*. In this framework, arbitrary fields act between pairs of "charged" particles. As long as the field satisfies a few simple conditions, namely that field lines start and end at the charges, that the flux is conserved and that a generalized superposition principle holds, it provably transports one data distribution to the other. Many concrete realizations of IFM are possible. The authors of (Manukhov et al., 2026) propose a field inspired by the strong interaction of subnuclear particles (quarks) in physics. This realization resolves the EFM issues discussed in §2.2.

From this perspective, C^{2}FM, along with PFGM and EFM, constitutes a particular realization of the IFM framework. However, C^{2}FM differs fundamentally from the strong-interaction-based realization of IFM in two respects. First, the IFM realization introduces the field of a *pair* of charges, $\mathbf{E}_{q\bar{q}}$, which cannot be decomposed as $\mathbf{E}_q + \mathbf{E}_{\bar{q}}$, whereas in electrostatics the superposition principle (8) holds. Second, the strong-interaction realization exhibits an explicit dependence on the transport plan $\pi(\cdot, \cdot)$, while our method is transport-plan agnostic.

For further context, we mention the work of Shlenskii et al. (2026), which uncovers a deep link between IFM and Conditional Flow Matching (CFM). It demonstrates that conventional CFM is a subset of IFM, termed *forward-only-IFM*, and that other IFM realizations, in particular EFM, are strictly more expressive than CFM.

5 Training and Inference

This section turns the construction into a method. First, in §5.1, we derive the training procedure; then, in §5.2, the sampler; finally, in §5.3, we validate both.

5.1 Training

The last difficulty in §2.2 is not geometric, and it has two parts. The EFM target is a normalised Monte-Carlo field estimate, and normalisation is nonlinear, so an unbiased $\widehat{\mathbf{E}}_M$ still yields a biased direction. Its training points come from a heuristic p_{vol} that the construction does not supply. We remove both by regressing conditional expectations instead of the direction: every target is computed from a single latent draw, and the query distribution is the Coulomb kernel of the field itself, which

we sample exactly. Detailed mathematical formulations, statements, and theorems for the training procedure are presented in Appendix D. The proofs of the statements are provided in Appendix E.

Field Decomposition. In $\mathbf{E}_C(\mathbf{x}, z)$, neither the query position $\mathbf{x}$ nor the auxiliary coordinate z is integrated over. They can therefore be separated from the Coulomb integrals, while all dependence on the source point $\mathbf{x}'$ is absorbed into weighted averages over the corresponding plate distribution, which completely determine the local field direction.

Theorem 5.1 (Denoiser representation of the field)**.** *Let $\mathbb{P}_0 = \mathbb{P}$, $\mathbb{P}_1 = \mathbb{Q}$, $z_n = nL$, and $K_{i,n} = \big[\|\mathbf{x} - \mathbf{x}'\|^2 + (z - z_{2n+i})^2\big]^{-(D+1)/2}$. Define, for $i = 0, 1$,*

$$\begin{aligned} \lambda_{i,C} &= \sum_{n=-C}^{C} \int K_{i,n}\, d\mathbb{P}_i, & \overline{\mathbf{x}}_i &= \frac{1}{\lambda_{i,C}} \sum_{n=-C}^{C} \int \mathbf{x}' K_{i,n}\, d\mathbb{P}_i, \\ \overline{h}_i &= \frac{(-1)^i}{\lambda_{i,C}} \sum_{n=-C}^{C} \int (z - z_{2n+i}) K_{i,n}\, d\mathbb{P}_i, & \ell &= \log \frac{\lambda_{0,C}}{\lambda_{1,C}}. \end{aligned} \tag{10}$$

Then, with $a = \sigma(\ell)$, the direction of the electric field is proportional to

$$\mathbf{E}_C \propto \Big(a\big(\mathbf{x} - \overline{\mathbf{x}}_0\big) + (1-a)\big(\overline{\mathbf{x}}_1 - \mathbf{x}\big),\;\; a\,\overline{h}_0 + (1-a)\,\overline{h}_1 \Big). \tag{11}$$

Each quantity has a direct geometric meaning. $\overline{\mathbf{x}}_i$ is the Coulomb-weighted effective source location of side i in the data coordinates, as seen from the current query point. $\overline{h}_i$ measures the net axial contribution of that side after the fields of its periodic copies partially cancel. Finally, ℓ controls the relative weight of the two sides in the total field. Thus, learning the field direction reduces to learning these five effective quantities.

New Objectives. We now construct training targets for the quantities entering Theorem 5.1. The key idea is to sample query points from the same Coulomb kernel that defines the field itself, rather than from an externally chosen training volume. We draw $\mathbf{x}_0 \sim \mathbb{P}$, $\mathbf{x}_1 \sim \mathbb{Q}$, and $z \sim p_z$ on $(0, L)$. We then choose a side i uniformly, sample one of its copies n according to Eq. (64), and generate

$$\mathbf{x} = \mathbf{x}_i + |z - z_{2n+i}|\, \frac{\boldsymbol{\xi}}{|\xi_z|}, \qquad \boldsymbol{\xi} \sim \mathcal{N}(0, I_D), \quad \xi_z \sim \mathcal{N}(0, 1). \tag{12}$$

Proposition D.2 shows that this procedure samples $\mathbf{x}$ from the normalized copy-summed Coulomb kernel $q_i(\mathbf{x}|\mathbf{x}_i, z)$ appearing in Eq. (10), up to a factor depending only on z (see (66)). Thus, unlike EFM, no separate heuristic distribution p_{vol} over the augmented space is required: the geometry of the field itself determines how training points are proposed. The remaining choice is only the one-dimensional distribution p_z, and the consistency statements below hold for any p_z with support throughout $(0, L)$.

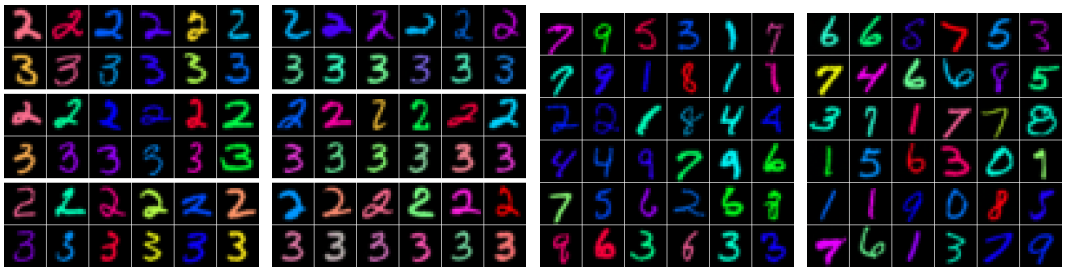
Because either side can generate the same query point, the two proposals are combined into the mixture $q_{\mathrm{mix}} = \frac{1}{2}(q_0 + q_1)$. Their contributions are separated by the importance weights $\gamma_i = q_i / q_{\mathrm{mix}}$, given explicitly in Eq. (68). These weights undo the proposal mixture and allow each required field quantity to be recovered from a target computed from a single latent sample.

Conditional means. The effective source locations $\overline{\mathbf{x}}_0$ and $\overline{\mathbf{x}}_1$ are obtained by ordinary weighted denoising. For any $\kappa_i(\mathbf{x}, z) > 0$ independent of the latent sample,

$$\mathcal{L}_{\mathrm{den}} = \mathbb{E}\Big[\gamma_0 \kappa_0 \big\|\overline{\mathbf{x}}_{\theta,0} - \mathbf{x}_0\big\|^2 + \gamma_1 \kappa_1 \big\|\overline{\mathbf{x}}_{\theta,1} - \mathbf{x}_1\big\|^2\Big] \tag{13}$$

is minimised by $\overline{\mathbf{x}}_{\theta,i} = \overline{\mathbf{x}}_i$. Thus the latent source point $\mathbf{x}_i$ itself is the single-sample target, while the Coulomb sampling and the weights γ_i turn the usual conditional mean into precisely the Coulomb-weighted mean required by the field representation. In practice, κ_i is chosen to balance different noise levels using the residual scale described in Eqs. (72) and (75).

Axial quantities. The axial quantities require a different target because they describe the partial cancellation between different copies of the same plate. We first factor out a geometry-based baseline $\widetilde{h}_i$, defined in Eq. (78), and let the network predict only a positive multiplicative correction, $\overline{h}_{\theta,i} = \widetilde{h}_i\, e^{g_{\theta,i}}$. For a known latent source sample $\mathbf{x}_i$, its exact contribution $h_i(\mathbf{x}_i; \mathbf{x}, z)$ is available from Proposition D.5. Hence a single sample provides the correction target $y_i = h_i(\mathbf{x}_i; \mathbf{x}, z)/\widetilde{h}_i$.



(a) C^2FM (FID=0.81) (b) FM (FID=16.85) (c) C^2FM (FID=0.74) (d) FM (FID=0.48)

Figure 4: Qualitative and quantitative results of our C^2FM on colored MNIST $2 \to 3$ translation **(a, b)** and noise $\to$ colored MNIST generation **(c, d)**, compared to Flow Matching (FM). In **(a, b)** rows alternate a source digit and its image. FID is computed on 10 000 samples.

With the Bregman divergence (Banerjee et al., 2005) of the exponential,

$$\mathcal{L}_{\mathrm{ax}} = \mathbb{E}\big[\gamma_0 D_{\mathrm{B}}(y_0, g_{\theta,0}) + \gamma_1 D_{\mathrm{B}}(y_1, g_{\theta,1})\big], \qquad D_{\mathrm{B}}(y, g) = e^g - yg + y\log y - y, \tag{14}$$

is minimised by $e^{g^*_{\theta,i}} = \mathbb{E}[y_i \mid \mathbf{x}, z]$, and therefore $\overline{h}^*_{\theta,i} = \overline{h}_i$. The baseline only removes the dominant geometric dependence; the exact Coulomb-weighted axial quantity is still fixed by the population optimum.

Mixing. Finally, the relative strength of the two sides can also be learned from the same samples. After applying the weights γ_i, the side index is an equally probable binary label. Let m_i denote the marginal density of a query proposed by side i, and define $\rho_i(z) = \sum_{n=-C}^{C} |z - z_{2n+i}|^{-1}$. The exact mixing logit then separates into $\ell = \log(m_0/m_1) + \log(\rho_0/\rho_1)$. The second term is fixed entirely by the capacitor geometry, so the network only needs to learn the density-ratio term $b_\theta \approx \log(m_0/m_1)$:

$$\mathcal{L}_{\mathrm{mix}} = \mathbb{E}\big[\mathrm{BCE}\big(b_\theta(\mathbf{x}, z);\, \mathbb{1}[i=0]\big)\big] \tag{15}$$

is minimised by $b^*_\theta = \log(m_0/m_1)$, hence $\ell^*_\theta = \ell$. For the experiments, we additionally use the closed-form approximation $\widetilde{\ell}$ derived in Eq. (88), avoiding the need to train this head separately.

Total. The three objectives therefore learn exactly the three ingredients needed to reconstruct the field direction: the effective source locations, the axial contributions, and their relative mixing weight. We combine them as

$$\mathcal{L} = \mathcal{L}_{\mathrm{den}} + \lambda_{\mathrm{ax}}\mathcal{L}_{\mathrm{ax}} + \lambda_{\mathrm{mix}}\mathcal{L}_{\mathrm{mix}}, \qquad \lambda_{\mathrm{ax}}, \lambda_{\mathrm{mix}} > 0. \tag{16}$$

Each term uses a target available from a single latent draw, while the query distribution itself is supplied by the Coulomb construction. The complete procedure is summarized in Algorithm 1. Details of the sampling weights, the choice of κ_i, the axial baseline, and the closed-form approximation used for the mixing logit are given in Appendix D.

5.2 Inference

At a point $(\mathbf{x}, z)$ the network supplies $\overline{\mathbf{x}}_{\theta,0}$, $\overline{\mathbf{x}}_{\theta,1}$, $\overline{h}_{\theta,0}$, $\overline{h}_{\theta,1}$ and ℓ_θ, and Theorem 5.1 turns them into the field direction. What remains is to follow it.

The axial clock is valid here, but still unusable. Electrostatic generative models integrate in the augmented coordinate: PFGM and EFM normalise the field by its axial component and solve

$$\frac{d\mathbf{x}}{dz} = \frac{\mathbf{E}_x(\mathbf{x}, z)}{E_z(\mathbf{x}, z)} \tag{17}$$

from one plate to the other (Xu et al., 2022; 2023; Kolesov et al., 2025). For C^2FM this is legitimate: by Theorem 3.1 $E_z > 0$ throughout the slab, so z is a valid parameter along every forward line. It is nonetheless not usable, for a structural reason. The chain confines lines by cancelling the axial push where the plates carry no mass, that cancellation is exactly what stops a line crossing a plate beside the atoms, and the cancelled quantity is the denominator of (17). Where the cancellation bites, E_z is driven towards zero while $\mathbf{E}_x$ is not: the line travels almost sideways, the right-hand side of (17) grows without bound, and no grid in z follows it, since the admissible step then varies by orders of magnitude along a single line (Appendix F). Confinement and the axial clock are the same quantity pulling in opposite directions.

Arc length. A field line is fixed by a direction and not by a speed, so the curve is unchanged by any positive reparametrisation and we may choose one that is well conditioned. We take the curve's own arc length,

$$\frac{d(\mathbf{x}, z)}{d\tau} = \frac{\mathbf{E}}{\|\mathbf{E}\|}. \tag{18}$$

Now there is no denominator to vanish: the right-hand side is a unit vector wherever the field is nonzero. Only the direction is ever formed, so the overall scale of $\mathbf{E}$ cancels, which matters, since the kernel carries $\|\cdot\|^{-(D+1)}$ and its magnitude underflows long before its direction stops being informative. The axial coordinate stays part of the state and stays monotone by Theorem 3.1; we have stopped driving by it, not tracking it.

Step size, and confinement of the discrete flow. Unit speed does not say how far to step. A Coulomb direction turns on the scale of the distance to the charge dominating it, and the assembly already reports where that charge is: at $\overline{\mathbf{x}}_{\theta,i}$ on plate i. With

$$d_i = \sqrt{\mathrm{gap}_i^2 + \|\mathbf{x} - \overline{\mathbf{x}}_{\theta,i}\|^2}, \quad \mathrm{gap}_0 = z, \;\; \mathrm{gap}_1 = L - z, \qquad h = \eta \, \min(d_0, d_1), \tag{19}$$

the step is a fixed fraction of that scale. It costs nothing: both means are already computed for the direction, and η is dimensionless. On a ray towards a plate a step ηd_i advances z by about ηz, so the sampler approaches a plate geometrically, at the same scale-free spacing the training coverage uses. We additionally cap h so that one step changes z by at most half the distance to the nearer plate. A line may then approach a plate arbitrarily closely but never step through it, and the discrete flow inherits the confinement of Theorem 3.1.

Launch and stopping. A trajectory is integrated over $0 < z_{\mathrm{start}} < z_{\mathrm{stop}} < L$. It starts at $\mathbf{x} = \mathbf{x}_0 + z_{\mathrm{start}} \, \boldsymbol{\xi}/|\xi_z|$ with $\mathbf{x}_0 \sim \mathbb{P}$: the same draw as (12), and the exact flux-weighted law of the points at which the field lines of an isolated source atom cross the plane $z = z_{\mathrm{start}}$. Integration proceeds by Heun's method in τ and stops when z reaches z_{stop}; the endpoint is the generated sample. The two offsets play different roles and are not symmetric: z_{start} is bounded below by what the network can resolve at the source plate and above by the spread of the launch, while $L - z_{\mathrm{stop}}$ is a numerical tolerance only. Their choice, and the value of η, are discussed in Appendix F. The full procedure is Algorithm 2.

5.3 Empirical Validation

To validate our algorithms we take the translation and generation tasks of §3.3 and apply Algorithms 1 and 2 to them. As a reference we use Flow Matching (Lipman et al., 2023, FM). For both models we report FID computed on $10\,000$ generated samples, together with the generations themselves. All additional technical details are provided in Appendix G, and the sampler settings in Appendix F.

Figure 4 shows that both algorithms are valid: C^{2}FM is far ahead of FM on translation and somewhat behind it on generation. Both FIDs are computed against the training split from the best checkpoint of each run (§G), so they compare the two models under identical settings rather than to numbers reported elsewhere.

6 Discussion and Limitations

We proposed C^{2}FM, a capacitor-chain field matching framework for data-to-data generative modeling. It keeps the electrostatic structure of EFM while removing its main geometric difficulties: flux leakage, backward-oriented field lines and the resulting stochastic branching. Moreover, it is trained by a single-sample objective, with no multi-sample field estimate anywhere.

Our experiments do not establish how the method scales: beyond two dimensions they cover a single dataset at a single resolution. Training is also expensive: the generation run had not converged within our iteration budget. We attribute part of this to the width of electrostatic field lines: by Proposition D.2 the query kernel is a mixture of multivariate Cauchy densities, one per plate copy. Raising the dimensionality of the augmented space above the single dimension used here replaces them by Student-t kernels and tempers the tail (Xu et al., 2023); that is a natural next step.

AI USE STATEMENT

AI tools assisted with polishing the text and checking proofs. All AI-assisted work was reviewed by the authors, who take responsibility for the final content of this work.

REPRODUCIBILITY STATEMENT

All assumptions behind our theoretical claims are stated explicitly and all proofs are given in Appendices A, C and E. Full details of our experiments are given in Appendices D, F and G, and the code is included in the supplementary materials.

REFERENCES

Michael Albergo, Nicholas M Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *Journal of Machine Learning Research*, 26(209):1–80, 2025.

Michael Samuel Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *The Eleventh International Conference on Learning Representations*, 2023.

Sheldon Axler, Paul Bourdon, and Wade Ramey. *Harmonic Function Theory*, volume 137 of *Graduate Texts in Mathematics*. Springer, New York, 2 edition, 2001. ISBN 978-0-387-95218-5. doi: 10.1007/978-1-4757-8137-3.

Arindam Banerjee, Srujana Merugu, Inderjit S. Dhillon, and Joydeep Ghosh. Clustering with bregman divergences. *Journal of Machine Learning Research*, 6(58):1705–1749, 2005.

Francisco Caruso, Vitor Oguri, and Felipe Silveira. Still learning about space dimensionality: From the description of hydrogen atom by a generalized wave equation for dimensions $d \geq 3$. *American Journal of Physics*, 91(2):153–158, 2023.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in neural information processing systems*, 35:26565–26577, 2022.

Alexander Kolesov, Stepan Manukhov, Vladimir V Palyulin, and Alexander Korotin. Field matching: an electrostatic paradigm to generate and transfer data. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=9dHilxylvC.

L. D. Landau and E. M. Lifshitz. *The Classical Theory of Fields (Volume 2 of A Course of Theoretical Physics)*. Pergamon Press, 1971.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022.

S. I. Manukhov, Alexander Kolesov, V.V. Palyulin, and Alexander Korotin. Interaction field matching: Overcoming limitations of electrostatic models. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=GEsTLuJy1q.

Daniil Shlenskii and Alexander Korotin. Overclocking electrostatic generative models. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=flo449mncA.

Daniil Shlenskii, Alexander Varlamov, Nazar Buzun, and Alexander Korotin. Unlocking the duality between flow and field matching. In *ICLR 2026 2nd Workshop on Deep Generative Model in Machine Learning: Theory, Principle and Efficacy*, 2026. URL https://openreview.net/forum?id=3Vw8H5Scsw.

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

Gerald Teschl. *Ordinary Differential Equations and Dynamical Systems*, volume 140 of *Graduate Studies in Mathematics*. American Mathematical Society, 2012. URL https://www.mat.univie.ac.at/~gerald/ftp/book-ode/ode.pdf.

Yilun Xu, Ziming Liu, Max Tegmark, and Tommi Jaakkola. Poisson flow generative models. *Advances in Neural Information Processing Systems*, 35:16782–16795, 2022.

Yilun Xu, Ziming Liu, Yonglong Tian, Shangyuan Tong, Max Tegmark, and Tommi Jaakkola. Pfgm++: Unlocking the potential of physics-inspired generative models. In *International Conference on Machine Learning*, pp. 38566–38591. PMLR, 2023.

A Proof of the Main Theorem

Theorem A.1 (Only forward-oriented field lines)**.** *Let a forward-oriented field line $d\widetilde{\mathbf{x}} = \mathbf{E}(\widetilde{\mathbf{x}})\, d\tau$ start at a point $\widetilde{\mathbf{x}} = (\mathbf{x}, z)$ with $\mathbf{x} \in \operatorname{supp} \mathbb{P}$, $z \to 0^+$, where $\mathbf{E}(\widetilde{\mathbf{x}})$ is the total electric field from Eq.* (8)*. Then this field line remains in the slab $z \in [0, L]$ and satisfies*

$$E_z(\mathbf{x}, z) > 0, \qquad 0 < z < L.$$

Proof. We first establish the reflection symmetry of the axial field. Fix a plane

$$z_0 = n_0 L, \qquad n_0 \in \mathbb{Z},$$

and let $\epsilon > 0$.

For the positive plates,

$$E_z^+(\mathbf{x}, z_0 - \epsilon) = \frac{1}{S_D} \sum_{n=-\infty}^{\infty} \int \frac{(n_0 - 2n)L - \epsilon}{[\|\mathbf{x} - \mathbf{x}'\|^2 + ((n_0 - 2n)L - \epsilon)^2]^{\frac{D+1}{2}}} \mathbb{P}(\mathbf{x}')\, d^D\mathbf{x}'. \tag{20}$$

Changing the summation index to

$$m = n_0 - n$$

gives

$$(n_0 - 2n)L - \epsilon = -\big((n_0 - 2m)L + \epsilon\big),$$

while the denominator remains unchanged. Therefore

$$E_z^+(\mathbf{x}, z_0 - \epsilon) = -E_z^+(\mathbf{x}, z_0 + \epsilon). \tag{21}$$

For the negative plates the same argument, with

$$m = n_0 - n - 1,$$

gives

$$E_z^-(\mathbf{x}, z_0 - \epsilon) = -E_z^-(\mathbf{x}, z_0 + \epsilon). \tag{22}$$

Hence the total field satisfies

$$E_z(\mathbf{x}, z_0 - \epsilon) = -E_z(\mathbf{x}, z_0 + \epsilon). \tag{23}$$

Importantly, Eq. (23) relates the two one-sided limits. On a charged plane the normal component of the electric field is discontinuous, so one should not identify either one-sided value with a single value $E_z(\mathbf{x}, z_0)$ on the sheet.

We now determine the field on the inner sides of the two plates. At the positive plate $z = 0$, Gauss's law gives

$$E_z(\mathbf{x}, 0^+) - E_z(\mathbf{x}, 0^-) = \mathbb{P}(\mathbf{x}). \tag{24}$$

Using the reflection relation

$$E_z(\mathbf{x}, 0^-) = -E_z(\mathbf{x}, 0^+),$$

we obtain

$$E_z(\mathbf{x}, 0^+) = \frac{1}{2}\mathbb{P}(\mathbf{x}) \geq 0. \tag{25}$$

Similarly, the surface charge on the plate $z = L$ is $-\mathbb{Q}(\mathbf{x})$, so

$$E_z(\mathbf{x}, L^+) - E_z(\mathbf{x}, L^-) = -\mathbb{Q}(\mathbf{x}). \tag{26}$$

Together with

$$E_z(\mathbf{x}, L^+) = -E_z(\mathbf{x}, L^-),$$

this gives

$$E_z(\mathbf{x}, L^-) = \frac{1}{2}\mathbb{Q}(\mathbf{x}) \geq 0. \tag{27}$$

We next prove positivity inside the slab. Since there are no charges in the open slab

$$\Omega = \mathbb{R}^D \times (0, L),$$

the electrostatic potential is harmonic,

$$\Delta\varphi = 0,$$

and therefore

$$\Delta E_z = -\partial_z \Delta\varphi = 0.$$

Thus, E_z is harmonic in Ω.

Compact support of the plate distributions together with the uniformly convergent capacitor-chain representation implies

$$E_z(\mathbf{x}, z) \to 0 \qquad \text{as} \qquad \|\mathbf{x}\| \to \infty.$$

The maximum principle for harmonic functions on possibly unbounded domains (Axler et al., 2001, Corollary 1.10) therefore gives

$$E_z(\mathbf{x}, z) \geq 0, \qquad (\mathbf{x}, z) \in \Omega.$$

If E_z vanished at an interior point, it would attain its minimum there. By the strong minimum principle (Axler et al., 2001, Theorem 1.8), E_z would then be constant throughout the connected domain Ω. This contradicts

$$E_z(\mathbf{x}, 0^+) = \frac{1}{2}\mathbb{P}(\mathbf{x}) > 0$$

on the support of $\mathbb{P}$. Hence

$$E_z(\mathbf{x}, z) > 0, \qquad \mathbf{x} \in \mathbb{R}^D, \quad 0 < z < L.$$

□

Theorem A.2 (C^{2}FM transport theorem). *Let $\mathbb{P}(\mathbf{x}^+)$ and $\mathbb{Q}(\mathbf{x}^-)$ be two data distributions with compact support, and let T be the deterministic map defined above. Let $\mathbf{x}^+$ be distributed as $\mathbb{P}(\mathbf{x}^+)$. Then $\mathbf{x}^- = T(\mathbf{x}^+)$ is distributed as $\mathbb{Q}(\mathbf{x}^-)$ almost surely:*

$$\textit{If } \mathbf{x}^+ \sim \mathbb{P}(\mathbf{x}^+) \Rightarrow T(\mathbf{x}^+) = \mathbf{x}^- \sim \mathbb{Q}(\mathbf{x}^-). \tag{28}$$

Proof. We verify that the field satisfies the interaction field requirements of (Manukhov et al., 2026, IFM). This automatically guarantees transport between the distributions.

- The first requirement of an interaction field is that a field line starting on a positive charge must almost surely (except, perhaps, a set of lines with zero flux) end on a negative charge. Because we work with electrostatic field lines, this requirement is fulfilled (see, e.g., (Kolesov et al., 2025, Appendix A)).
- The second requirement is flux conservation along a field-line tube between the planes $z \in (0, L)$. In this interval our charge configuration satisfies the requirement by Gauss's theorem (see the discussion in Kolesov et al. (2025)).
- Finally, the generalized superposition principle reduces, as shown in Manukhov et al. (2026), to the ordinary superposition principle (8) in the case of electrostatic fields.

Thus all IFM field requirements are met, which guarantees transport between the distributions. □

B Finite Capacitor Approximation

In practice, the infinite sums in (6) must be truncated to $|n| \leq C$, retaining $2C + 1$ capacitor cells. We give two theoretical bounds on this cutoff: the first controls the field itself, whereas the second controls the induced transport. Let

$$R = \sup_{\substack{\mathbf{p}\in\operatorname{supp}\mathbb{P} \\ \mathbf{q}\in\operatorname{supp}\mathbb{Q}}} \sqrt{\|\mathbf{p}-\mathbf{q}\|^2 + L^2}, \qquad A_D = \frac{R}{S_D L^{D+1}}. \tag{29}$$

Theorem B.1 (Field truncation). *Let $\mathbb{P}, \mathbb{Q}$ have compact support and let $\mathbf{E}_C := \sum_{n=-C}^{C}(\mathbf{E}_n^+ + \mathbf{E}_n^-)$. For any $\varepsilon_{\mathrm{fld}} > 0$,*

$$C \geq \left\lceil \frac{1}{2}\left[1 + \left(\frac{A_D}{\varepsilon_{\mathrm{fld}}}\right)^{1/D}\right]\right\rceil \quad \Longrightarrow \quad \sup_{\mathbf{x}\in\mathbb{R}^D,\, 0<z<L} \|\mathbf{E}_C - \mathbf{E}\| \leq \varepsilon_{\mathrm{fld}}. \tag{30}$$

Small field error alone does not imply small endpoint error: nearby vector fields may generate strongly diverging integral curves. To control transport, let $\mathbf{x}_\infty(z)$ be an ideal field line and introduce its a-tube $\mathcal{T}_a = \{(z, \mathbf{x}) : 0 \leq z \leq L,\ \|\mathbf{x} - \mathbf{x}_\infty(z)\| \leq a\}$. Let us define on this tube

$$m := \inf_{\mathcal{T}_a} E_z > 0, \qquad M := \sup_{\mathcal{T}_a} \left\| \frac{E_x}{E_z} \right\|$$
$$\Lambda := \sup_{\mathcal{T}_a} \left\| \nabla_{\mathbf{x}} \frac{E_x}{E_z} \right\| \tag{31}$$

Here m is the forward-field margin, M bounds the line slope, and Λ measures sensitivity to transverse perturbations.

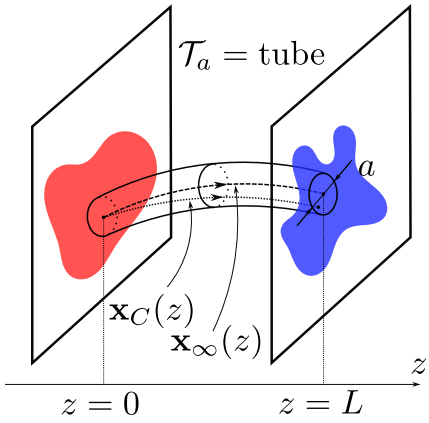


Figure 5: The finite-chain trajectory $\mathbf{x}_C(z)$ must remain inside the radius-a tube $\mathcal{T}_a$ around the ideal infinite-chain trajectory $\mathbf{x}_\infty(z)$.

Theorem B.2 (Finite-chain transport stability). *Let the ideal and truncated lines start at the same point, and let $\varepsilon_{\mathrm{tr}} > 0$. Define*

$$B := \frac{2(1+M)}{m},$$
$$H_{\Lambda,L} := \begin{cases} (e^{\Lambda L} - 1)/\Lambda, & \Lambda > 0, \\ L, & \Lambda = 0, \end{cases} \tag{32}$$
$$\zeta := \min\{\varepsilon_{\mathrm{tr}}, a/2\}.$$

Then $\|\mathbf{x}_C(L) - \mathbf{x}_\infty(L)\| \leq \varepsilon_{\mathrm{tr}}$ and $\mathbf{x}_C$ remains in $\mathcal{T}_a$ whenever

$$C \geq \left\lceil \frac{1}{2}\left[1 + \left(\max\left\{\frac{2A_D}{m}, \frac{A_D B H_{\Lambda,L}}{\zeta}\right\}\right)^{1/D}\right]\right\rceil. \tag{33}$$

Proofs of both statements are given in §C.

C Proof of the Finite-Chain Bounds

Proof of Theorem B.1. We estimate the field of each omitted capacitor and then sum these estimates. The positive and negative plates must be considered together: their charges cancel, making the field decay faster than the field of either plate separately.

The field of the n-th capacitor is

$$\mathbf{E}_n^+(\mathbf{x}, z) + \mathbf{E}_n^-(\mathbf{x}, z) = \frac{1}{S_D}\int \frac{(\mathbf{x} - \mathbf{p}, z - 2nL)}{\left(\|\mathbf{x} - \mathbf{p}\|^2 + (z - 2nL)^2\right)^{(D+1)/2}}\, d\mathbb{P}(\mathbf{p})$$
$$- \frac{1}{S_D}\int \frac{(\mathbf{x} - \mathbf{q}, z - (2n+1)L)}{\left(\|\mathbf{x} - \mathbf{q}\|^2 + (z - (2n+1)L)^2\right)^{(D+1)/2}}\, d\mathbb{Q}(\mathbf{q}). \tag{34}$$

Since both distributions are normalized to unity, we can combine these integrals:

$$\mathbf{E}_n^+ + \mathbf{E}_n^- = \frac{1}{S_D}\iint \left[\frac{\mathbf{r}_+}{\|\mathbf{r}_+\|^{D+1}} - \frac{\mathbf{r}_-}{\|\mathbf{r}_-\|^{D+1}}\right] d\mathbb{P}(\mathbf{p})\, d\mathbb{Q}(\mathbf{q}), \tag{35}$$

where

$$\mathbf{r}_+ = (\mathbf{x} - \mathbf{p}, z - 2nL), \qquad \mathbf{r}_- = (\mathbf{x} - \mathbf{q}, z - (2n+1)L)$$

are the vectors from the positive and negative charges to the observation point. Their difference satisfies

$$\|\mathbf{r}_- - \mathbf{r}_+\| = \sqrt{\|\mathbf{p}-\mathbf{q}\|^2 + L^2} \le R.$$

We now estimate the difference in square brackets. Join $\mathbf{r}_+$ and $\mathbf{r}_-$ by the straight segment

$$\mathbf{r}(t) = (1-t)\mathbf{r}_+ + t\mathbf{r}_-, \qquad 0 \le t \le 1.$$

For an omitted capacitor, $|n| > C \ge 1$, every point on this segment has

$$\|\mathbf{r}(t)\| \ge |z - (2n+t)L| \ge (2|n|-1)L, \qquad 0 < z < L. \tag{36}$$

Indeed, for positive n the entire capacitor lies above $2nL$, while for negative n it lies below $(2n+1)L$. This distance estimate holds for every $\mathbf{x}$.

Also note that:

$$\|\mathbf{r}'(t)\| = \|\mathbf{r}_- - \mathbf{r}_+\| \le R \tag{37}$$

Differentiating the Coulomb field along the segment gives

$$\frac{d}{dt}\frac{\mathbf{r}(t)}{\|\mathbf{r}(t)\|^{D+1}} = \frac{1}{\|\mathbf{r}(t)\|^{D+1}}\left[\mathbf{r}'(t) - (D+1)\frac{\mathbf{r}(t)\cdot\mathbf{r}'(t)}{\|\mathbf{r}(t)\|^2}\mathbf{r}(t)\right]. \tag{38}$$

To bound this expression, resolve $\mathbf{r}'(t)$ into components parallel and perpendicular to $\mathbf{r}(t)$. The expression in square brackets becomes $\mathbf{r}'_\perp - D\mathbf{r}'_\parallel$. Therefore, for $D \ge 1$,

$$\|\mathbf{r}'_\perp - D\mathbf{r}'_\parallel\|^2 = \|\mathbf{r}'_\perp\|^2 + D^2\|\mathbf{r}'_\parallel\|^2 \le D^2\|\mathbf{r}'(t)\|^2 \le D^2R^2.$$

Together with (36), this yields

$$\left\|\frac{d}{dt}\frac{\mathbf{r}(t)}{\|\mathbf{r}(t)\|^{D+1}}\right\| \le \frac{DR}{[(2|n|-1)L]^{D+1}}. \tag{39}$$

Integrating from $t=0$ to $t=1$, we obtain

$$\left\|\frac{\mathbf{r}_+}{\|\mathbf{r}_+\|^{D+1}} - \frac{\mathbf{r}_-}{\|\mathbf{r}_-\|^{D+1}}\right\| \le \frac{DR}{[(2|n|-1)L]^{D+1}}. \tag{40}$$

Thus, the cancellation between the opposite charges gives a bound proportional to their separation R and to the inverse $(D+1)$-st power of the distance.

Substituting this estimate into (35) and using the normalization of $\mathbb{P}$ and $\mathbb{Q}$, we find

$$\|\mathbf{E}_n^+ + \mathbf{E}_n^-\| \le \frac{DR}{S_D L^{D+1}}\frac{1}{(2|n|-1)^{D+1}}. \tag{41}$$

We can now sum over all omitted capacitors:

$$\begin{aligned}\|\mathbf{E} - \mathbf{E}_C\| &\le \sum_{|n|>C}\|\mathbf{E}_n^+ + \mathbf{E}_n^-\| \\ &\le \frac{2DR}{S_D L^{D+1}}\sum_{n=C+1}^{\infty}\frac{1}{(2n-1)^{D+1}}.\end{aligned} \tag{42}$$

The factor 2 accounts for the omitted cells on both sides of the retained chain. Since the summand decreases with n,

$$\sum_{n=C+1}^{\infty}\frac{1}{(2n-1)^{D+1}} \le \int_C^{\infty}\frac{ds}{(2s-1)^{D+1}} = \frac{1}{2D(2C-1)^D}. \tag{43}$$

Consequently,

$$\|\mathbf{E}_C(\mathbf{x},z) - \mathbf{E}(\mathbf{x},z)\| \le \frac{R}{S_D L^{D+1}(2C-1)^D} = \frac{A_D}{(2C-1)^D}. \tag{44}$$

This bound is uniform for all $\mathbf{x} \in \mathbb{R}^D$ and $0 < z < L$. It also establishes absolute convergence of the series of capacitor fields.

Finally, requiring the right-hand side to be at most $\varepsilon_{\mathrm{fld}}$ gives

$$2C - 1 \geq \left(\frac{A_D}{\varepsilon_{\mathrm{fld}}}\right)^{1/D}.$$

Since C is an integer, it is sufficient to choose

$$C \geq \left\lceil \frac{1}{2}\left[1 + \left(\frac{A_D}{\varepsilon_{\mathrm{fld}}}\right)^{1/D}\right]\right\rceil,$$

which proves the theorem. □

Proof of Theorem B.2. We first show that truncation preserves the forward direction of the field inside the tube. We then estimate the change in the line slope, integrate the resulting trajectory error, and verify that the truncated line cannot leave the tube.

All field values at the plates are understood as one-sided limits from inside the slab. We use the regularity assumptions stated in the main text; in particular, the field-line equations are locally well posed near the tube.

Denote the omitted field by

$$\delta\mathbf{E} := \mathbf{E}_C - \mathbf{E}, \qquad \eta_C := A_D(2C-1)^{-D}. \tag{45}$$

The first theorem gives $\|\delta\mathbf{E}\| \leq \eta_C$ everywhere in the slab, and therefore on the tube. By continuity, the same estimate holds for the one-sided boundary values.

Preservation of forward motion. Since $E_z \geq m$ in $\mathcal{T}_a$, the truncated axial field obeys

$$E_{C,z} = E_z + \delta E_z \geq m - |\delta E_z| \geq m - \eta_C. \tag{46}$$

Thus the condition $\eta_C \leq m/2$ implies

$$E_{C,z} \geq \frac{m}{2} > 0 \qquad \text{throughout } \mathcal{T}_a. \tag{47}$$

The finite-chain line therefore also moves forward and can be parametrized by z. Its equation and the ideal equation are

$$\frac{d\mathbf{x}_C}{dz} = \frac{\mathbf{E}_{C,x}(\mathbf{x}_C(z),z)}{E_{C,z}(\mathbf{x}_C(z),z)}, \qquad \frac{d\mathbf{x}_\infty}{dz} = \frac{\mathbf{E}_x(\mathbf{x}_\infty(z),z)}{E_z(\mathbf{x}_\infty(z),z)}. \tag{48}$$

These ratios are the transverse displacements per unit axial displacement, rather than velocities in physical time.

Error in the line slope. First compare the two slopes at the same spatial point. Subtracting the ratios exactly gives

$$\begin{aligned}\frac{\mathbf{E}_{C,x}}{E_{C,z}} - \frac{\mathbf{E}_x}{E_z} &= \frac{\mathbf{E}_x + \delta\mathbf{E}_x}{E_z + \delta E_z} - \frac{\mathbf{E}_x}{E_z} \\ &= \frac{\delta\mathbf{E}_x - (\mathbf{E}_x/E_z)\delta E_z}{E_{C,z}}.\end{aligned} \tag{49}$$

The numerator contains two effects: a direct perturbation of the transverse field and a change in the axial field used to parametrize the line. Using $\|\mathbf{E}_x/E_z\| \leq M$ and (47), we obtain

$$\begin{aligned}\left\|\frac{\mathbf{E}_{C,x}}{E_{C,z}} - \frac{\mathbf{E}_x}{E_z}\right\| &\leq \frac{2}{m}\left(\|\delta\mathbf{E}_x\| + M|\delta E_z|\right) \\ &\leq \frac{2(1+M)}{m}\eta_C = B\eta_C.\end{aligned} \tag{50}$$

In particular, a small axial field makes the direction sensitive to an otherwise small absolute field error.

Accumulation along the trajectory. Let

$$e(z) := \|\mathbf{x}_C(z) - \mathbf{x}_\infty(z)\|.$$

As long as the truncated trajectory stays in the tube, the straight transverse segment between the two trajectories at a fixed z also belongs to the tube: its cross-section is a Euclidean ball.

Integrating the transverse derivative of the ideal slope along this segment therefore gives

$$\begin{aligned}\left\|\frac{\mathbf{E}_x}{E_z}(\mathbf{x}_C(z),z)-\frac{\mathbf{E}_x}{E_z}(\mathbf{x}_\infty(z),z)\right\| \\ \le \sup_{\mathcal{T}_a}\left\|\nabla_{\mathbf{x}}\frac{\mathbf{E}_x}{E_z}\right\|\|\mathbf{x}_C(z)-\mathbf{x}_\infty(z)\| = \Lambda e(z).\end{aligned} \tag{51}$$

Subtracting (48) and adding and subtracting the ideal slope at $\mathbf{x}_C(z)$, we obtain

$$\begin{aligned}\left\|\frac{d}{dz}(\mathbf{x}_C-\mathbf{x}_\infty)\right\| &= \left\|\frac{\mathbf{E}_{C,x}}{E_{C,z}}(\mathbf{x}_C,z)-\frac{\mathbf{E}_x}{E_z}(\mathbf{x}_\infty,z)\right\| \\ &= \left\|\left[\frac{\mathbf{E}_{C,x}}{E_{C,z}}(\mathbf{x}_C,z)-\frac{\mathbf{E}_x}{E_z}(\mathbf{x}_C,z)\right]+\left[\frac{\mathbf{E}_x}{E_z}(\mathbf{x}_C,z)-\frac{\mathbf{E}_x}{E_z}(\mathbf{x}_\infty,z)\right]\right\| \\ &\le B\eta_C+\Lambda e(z).\end{aligned} \tag{52}$$

The first term compares the two fields at the same point and is bounded by (50). The second compares the ideal field slope at two different points and is bounded by (51).

The trajectories start at the same point, so $e(0)=0$. The norm e is absolutely continuous, and its derivative satisfies, almost everywhere,

$$e'(z)\le B\eta_C+\Lambda e(z). \tag{53}$$

This is the differential form of the Gronwall estimate (Teschl, 2012, Lemma 2.7). For completeness, multiplication by the integrating factor $e^{-\Lambda z}$ gives

$$\frac{d}{dz}\big(e^{-\Lambda z}e(z)\big)\le B\eta_C e^{-\Lambda z}.$$

Integration from 0 to z then yields

$$e(z)\le B\eta_C\int_0^z e^{\Lambda(z-s)}\,ds = B\eta_C\begin{cases}(e^{\Lambda z}-1)/\Lambda, & \Lambda>0,\\ z, & \Lambda=0.\end{cases} \tag{54}$$

The exponential factor bounds how an error introduced at axial position s can grow before reaching z. In particular,

$$e(z)\le BH_{\Lambda,L}\eta_C,\qquad 0\le z\le L, \tag{55}$$

as long as the trajectory remains in the tube.

Self-consistency of the tube estimate. The bound (55) was derived only while the truncated trajectory remains inside $\mathcal{T}_a$, where our assumptions on the field hold. Before applying it at $z=L$, we must therefore show that the trajectory cannot leave this tube or terminate before reaching the target plate. Suppose that

$$\eta_C\le\frac{m}{2},\qquad BH_{\Lambda,L}\eta_C\le\zeta=\min\{\varepsilon_{\mathrm{tr}},a/2\}. \tag{56}$$

If the truncated trajectory reached the tube boundary for the first time at some z_*, then $e(z_*)=a$. However, the estimate valid up to that first contact, together with continuity, gives $e(z_*)\le a/2$. This is a contradiction.

The solution also cannot terminate at an interior value of z without reaching the boundary: it stays in a compact region where the field-line equation is regular and its axial denominator is bounded away from zero. The standard ODE continuation criterion therefore extends it through every interior point (Teschl, 2012, Section 2.6). Moreover, its slope is bounded by $M+B\eta_C$, so it has a finite limit as $z\uparrow L$. Consequently, the truncated trajectory reaches the target plate without leaving $\mathcal{T}_a$, and

$$\|\mathbf{x}_C(L)-\mathbf{x}_\infty(L)\|\le BH_{\Lambda,L}\eta_C. \tag{57}$$

To guarantee the prescribed endpoint accuracy, we therefore require $BH_{\Lambda,L}\eta_C\le\varepsilon_{\mathrm{tr}}$. Together with the conditions ensuring forward motion and confinement to the tube, this gives three simultaneous requirements:

$$\eta_C\le\frac{m}{2},\qquad BH_{\Lambda,L}\eta_C\le\frac{a}{2},\qquad BH_{\Lambda,L}\eta_C\le\varepsilon_{\mathrm{tr}}. \tag{58}$$

Substituting $\eta_C = A_D(2C-1)^{-D}$, we obtain

$$(2C-1)^D \geq \max\left\{\frac{2A_D}{m}, \frac{2A_D B H_{\Lambda,L}}{a}, \frac{A_D B H_{\Lambda,L}}{\varepsilon_{\mathrm{tr}}}\right\}. \tag{59}$$

The maximum ensures that all three requirements hold. Using $\zeta = \min\{\varepsilon_{\mathrm{tr}}, a/2\}$, the last two entries combine into $A_D B H_{\Lambda,L}/\zeta$. Taking the D-th root and rounding the resulting lower bound on C upwards gives (33). □

D Learning Details

This appendix provides the formal definitions and technical details underlying the learning procedure described in §5.1. We first derive the representation of the capacitor-chain field in terms of Coulomb-weighted moments, and then construct the transverse and axial learning objectives together with the analytic approximation to the mixing logit. Proofs of the corresponding statements are collected in Appendix E.

D.1 Field representation through weighted moments

Consider the finite-chain field $\mathbf{E}_C(\mathbf{x}, z)$ obtained from Eq. (6) by restricting the sums to $|n| \leq C$. Introducing $z_n = nL$ we define, for $i = 0, 1$,

$$\lambda_{i,C}(\mathbf{x}, z) = \sum_{n=-C}^{C} \int \frac{\mathbb{P}_i(\mathbf{x}')\, d^D\mathbf{x}'}{(\|\mathbf{x}-\mathbf{x}'\|^2 + (z_{2n+i}-z)^2)^{\frac{D+1}{2}}}, \tag{60}$$

where $\mathbb{P}_0 = \mathbb{P}$ and $\mathbb{P}_1 = \mathbb{Q}$. The corresponding weighted means are

$$\begin{aligned} \overline{\mathbf{x}}_i &= \frac{1}{\lambda_{i,C}} \sum_{n=-C}^{C} \int \frac{\mathbf{x}'\mathbb{P}_i(\mathbf{x}')\, d^D\mathbf{x}'}{(\|\mathbf{x}-\mathbf{x}'\|^2 + (z_{2n+i}-z)^2)^{\frac{D+1}{2}}}, \\ \overline{h}_i &= \frac{(-1)^i}{\lambda_{i,C}} \sum_{n=-C}^{C} \int \frac{(z - z_{2n+i})\mathbb{P}_i(\mathbf{x}')\, d^D\mathbf{x}'}{(\|\mathbf{x}-\mathbf{x}'\|^2 + (z_{2n+i}-z)^2)^{\frac{D+1}{2}}}. \end{aligned} \tag{61}$$

Finally, we introduce

$$\ell(\mathbf{x}, z) = \log \frac{\lambda_{0,C}(\mathbf{x}, z)}{\lambda_{1,C}(\mathbf{x}, z)}. \tag{62}$$

Theorem D.1 (Denoiser representation of the field)**.** *Let* $a(\mathbf{x}, z) = \sigma(\ell(\mathbf{x}, z))$[1]. *Then*

$$\mathbf{E}_C(\mathbf{x}, z) \propto \left(a(\mathbf{x} - \overline{\mathbf{x}}_0) + (1-a)(\overline{\mathbf{x}}_1 - \mathbf{x}),\ a\,\overline{h}_0 + (1-a)\,\overline{h}_1\right). \tag{63}$$

The proof is given in Appendix E in E.1. Hence, the field direction is completely determined by $\overline{\mathbf{x}}_0$, $\overline{\mathbf{x}}_1$, $\overline{h}_0$, $\overline{h}_1$, and ℓ.

D.2 Transverse denoising objective

We now construct a regression objective for the transverse moments $\overline{\mathbf{x}}_0$ and $\overline{\mathbf{x}}_1$. We first sample

$$\mathbf{x}_0 \sim \mathbb{P}, \qquad \mathbf{x}_1 \sim \mathbb{Q}, \qquad z \sim p_z(z),$$

where p_z is any coverage distribution with support in $(0, L)$. In our experiments, we use a symmetric log-uniform coverage in the distance to the nearest plate. Specifically, we sample $r \sim \mathrm{LogUniform}(\varepsilon, L/2)$ and set $z = r$ or $z = L - r$ with equal probability. This allocates more training samples to the regions close to the two plates while preserving coverage of the entire slab.

[1] where the standard logistic sigmoid $\sigma(u) = (1 + e^{-u})^{-1}$.

Next, a side $i \in \{0, 1\}$ is chosen with equal probability and a copy $n \in \{-C, \dots, C\}$ is sampled according to

$$\Pr(n \mid i, z) = \frac{|z - z_{2n+i}|^{-1}}{\sum_{m=-C}^{C} |z - z_{2m+i}|^{-1}}. \tag{64}$$

The training query is then drawn from the corresponding Coulomb kernel,

$$\mathbf{x} = \mathbf{x}_i + |z - z_{2n+i}| \frac{\boldsymbol{\xi}}{|\xi_z|}, \qquad \boldsymbol{\xi} \sim \mathcal{N}(0, I_D), \quad \xi_z \sim \mathcal{N}(0, 1). \tag{65}$$

The sampling procedure above has a direct electrostatic interpretation. The conditional density of $\mathbf{x}$ generated from a fixed $\mathbf{x}_i$ is the normalized Coulomb kernel.

Proposition D.2 (Coulomb-kernel sampling)**.** *For $i = 0, 1$, Eqs.* (64)–(12) *induce the conditional density*

$$q_i(\mathbf{x} \mid \mathbf{x}_i, z) = \frac{\Gamma\left(\frac{D+1}{2}\right)}{\pi^{\frac{D+1}{2}}} \frac{\sum_{n=-C}^{C} \left(\|\mathbf{x} - \mathbf{x}_i\|^2 + (z - z_{2n+i})^2\right)^{-\frac{D+1}{2}}}{\sum_{n=-C}^{C} |z - z_{2n+i}|^{-1}}. \tag{66}$$

Each copy contributes a multivariate Cauchy kernel, i.e. a Student-t with one degree of freedom, and q_i is their mixture.

Thus, up to a normalization factor depending only on z, the conditional density coincides with the Coulomb kernel entering Eq. (60). Since the training queries are sampled from an equal mixture of the two sides, we introduce

$$q_{\text{mix}}(\mathbf{x} \mid \mathbf{x}_0, \mathbf{x}_1, z) = \frac{1}{2} \left[q_0(\mathbf{x} \mid \mathbf{x}_0, z) + q_1(\mathbf{x} \mid \mathbf{x}_1, z) \right], \tag{67}$$

and

$$\gamma_i = \frac{q_i(\mathbf{x} \mid \mathbf{x}_i, z)}{q_{\text{mix}}(\mathbf{x} \mid \mathbf{x}_0, \mathbf{x}_1, z)}, \qquad i = 0, 1. \tag{68}$$

Theorem D.3 (Consistency of the joint denoising objective)**.** *Let $\kappa_i(\mathbf{x}, z) > 0$, $i = 0, 1$, be arbitrary functions independent of the latent samples $\mathbf{x}_i$. Then the population loss*

$$\begin{aligned} \mathcal{L}_{\text{den}} = \mathbb{E}_{\substack{z \sim p_z,\ \mathbf{x}_0 \sim \mathbb{P},\ \mathbf{x}_1 \sim \mathbb{Q}, \\ \mathbf{x} \sim q_{\text{mix}}(\cdot \mid \mathbf{x}_0, \mathbf{x}_1, z)}} \Big[& \gamma_0 \kappa_0 \left\| \overline{\mathbf{x}}_{\theta,0}(\mathbf{x}, z) - \mathbf{x}_0 \right\|^2 \\ & + \gamma_1 \kappa_1 \left\| \overline{\mathbf{x}}_{\theta,1}(\mathbf{x}, z) - \mathbf{x}_1 \right\|^2 \Big] \end{aligned} \tag{69}$$

is minimized pointwise by

$$\overline{\mathbf{x}}_{\theta,0}(\mathbf{x}, z) = \overline{\mathbf{x}}_0(\mathbf{x}, z), \qquad \overline{\mathbf{x}}_{\theta,1}(\mathbf{x}, z) = \overline{\mathbf{x}}_1(\mathbf{x}, z), \tag{70}$$

where the right-hand sides are the Coulomb-weighted means defined in Eq. (61)*.*

The proofs of Proposition D.2 and Theorem D.3 are given in Appendix E in E.2 and E.3.

As a technical improvement, we additionally balance the contribution of different noise levels. To prevent noisy regions with intrinsically larger denoising errors from dominating the objective, we introduce the local residual scale

$$c_{\text{out},i}^2(\mathbf{x}, z) = \frac{1}{D} \mathbb{E}\Big[\left\| \mathbf{x}_i - \overline{\mathbf{x}}_i(\mathbf{x}, z) \right\|^2 \Big| \, \mathbf{x}, z, i \Big]. \tag{71}$$

We then choose

$$\kappa_i(\mathbf{x}, z) = c_{\text{out},i}^{-2}(\mathbf{x}, z). \tag{72}$$

Proposition D.4 (Balanced denoising objective)**.** *With the choice* (72)*, the irreducible contribution of each query* $(\mathbf{x}, z)$ *to the denoising loss is normalized to the same scale:*

$$\frac{1}{D}\,\kappa_i\,\mathbb{E}\Big[\|\mathbf{x}_i-\overline{\mathbf{x}}_i\|^2\,\Big|\,\mathbf{x},z,i\Big]=1. \tag{73}$$

Thus, the network is trained on the displacement relative to its local uncertainty scale rather than being dominated by regions where this displacement is intrinsically large.

The proof is given in Appendix E in E.4.

The exact residual scale in Eq. (71) is not available during training. We therefore use a simple Gaussian approximation. Let

$$\omega_i(z)=\frac{r_i(z)}{\operatorname{med}|N(0,1)|}\simeq 1.48\,r_i(z),\qquad r_0(z)=z,\qquad r_1(z)=L-z. \tag{74}$$

Approximating the i-th plate distribution by an isotropic Gaussian with variance σ_i^2 per coordinate, we use

$$\widetilde{c}^{\,2}_{\mathrm{out},i}(z)=\frac{\sigma_i^2\omega_i^2(z)}{\sigma_i^2+\omega_i^2(z)},\qquad \kappa_i(z)=\widetilde{c}^{\,-2}_{\mathrm{out},i}(z). \tag{75}$$

This is the posterior variance per coordinate in the corresponding Gaussian denoising problem. Since $\kappa_i(z)>0$ is independent of the latent sample $\mathbf{x}_i$, Theorem D.3 guarantees that this practical reweighting does not change the population minimizer.

D.3 Axial denoising objective

The transverse denoisers $\overline{\mathbf{x}}_0$ and $\overline{\mathbf{x}}_1$ are not sufficient to reconstruct the full field. Indeed, the axial component also depends on the weighted means $\overline{h}_0$ and $\overline{h}_1$ introduced in Eq. (61). These quantities describe the compensation between different copies of the same plate and therefore cannot, in general, be inferred from the transverse mean alone.

For a fixed source sample $\mathbf{x}_i$, its contribution to the axial moment is obtained by evaluating the finite-chain weights at this source position as stated in the following proposition.

Proposition D.5 (Exact single-sample axial contribution)**.** *For a known source sample* $\mathbf{x}_i$ *and a query point* $(\mathbf{x}, z)$*, the exact axial contribution of the corresponding finite chain is*

$$h_i(\mathbf{x}_i;\mathbf{x},z)=\frac{\displaystyle\sum_{n=-C}^{C}(-1)^i(z-z_{2n+i})\left[\|\mathbf{x}-\mathbf{x}_i\|^2+(z-z_{2n+i})^2\right]^{-\frac{D+1}{2}}}{\displaystyle\sum_{n=-C}^{C}\left[\|\mathbf{x}-\mathbf{x}_i\|^2+(z-z_{2n+i})^2\right]^{-\frac{D+1}{2}}},\qquad i=0,1. \tag{76}$$

The proof follows directly from the finite-chain field and is given in Appendix E in E.5.

Directly learning $\overline{h}_i$ from scratch is possible, but unnecessary. The transverse denoiser already provides a good estimate of the position of the corresponding source, and therefore of its transverse distance from the query. We use this information to construct a simple baseline and let the network learn only the remaining correction.

Using the same Gaussian uncertainty estimate as in Eq. (75), we approximate the conditional mean squared transverse distance by

$$\widetilde{R}_i^{\,2}=\|\mathbf{x}-\overline{\mathbf{x}}_{\theta,i}\|^2+\Delta_i,\qquad \Delta_i=D\,\widetilde{c}^{\,2}_{\mathrm{out},i}=D\,\frac{\sigma_i^2\omega_i^2}{\sigma_i^2+\omega_i^2}. \tag{77}$$

Thus, Δ_i accounts for the residual uncertainty of the transverse source position that is not captured by its conditional mean $\overline{\mathbf{x}}_{\theta,i}$.

Substituting $\widetilde{R}_i$ into the finite-chain expression gives the baseline

$$\widetilde{h}_i = \frac{\sum_{n=-C}^{C} (-1)^i (z - z_{2n+i}) \left[\widetilde{R}_i^2 + (z - z_{2n+i})^2\right]^{-\frac{D+1}{2}}}{\sum_{n=-C}^{C} \left[\widetilde{R}_i^2 + (z - z_{2n+i})^2\right]^{-\frac{D+1}{2}}}. \tag{78}$$

Its purpose is only to capture the dominant geometric dependence of the axial moment. The remaining discrepancy is learned explicitly by the axial network, so an exact baseline is neither assumed nor required. We parameterize the final prediction as a multiplicative correction,

$$\overline{h}_{\theta,i} = \widetilde{h}_i \exp g_{\theta,i}(\mathbf{x}, z). \tag{79}$$

During training, the true source sample $\mathbf{x}_i$ is known, so Eq. (76) can be evaluated exactly. We define the corresponding correction target as

$$y_i = \frac{h_i(\mathbf{x}_i; \mathbf{x}, z)}{\widetilde{h}_i(\mathbf{x}, z)}. \tag{80}$$

To recover its conditional arithmetic mean, we use the Bregman divergence associated with the exponential function (Banerjee et al., 2005),

$$D_{\mathrm{B}}(y, g) = e^g - yg + y \log y - y, \tag{81}$$

and train the two axial heads with

$$\mathcal{L}_{\mathrm{ax}} = \mathbb{E}\left[\gamma_0 D_{\mathrm{B}}(y_0, g_{\theta,0}) + \gamma_1 D_{\mathrm{B}}(y_1, g_{\theta,1})\right]. \tag{82}$$

A standard property of Bregman divergences is that their population minimizer is the corresponding conditional mean (Banerjee et al., 2005). Therefore,

$$e^{g^*_{\theta,i}} = \mathbb{E}[y_i \mid \mathbf{x}, z], \tag{83}$$

and consequently

$$\overline{h}^*_{\theta,i} = \widetilde{h}_i \mathbb{E}\left[\frac{h_i}{\widetilde{h}_i}\middle| \mathbf{x}, z\right] = \overline{h}_i. \tag{84}$$

D.4 Analytic approximation of the mixing logit

The final quantity required for the field reconstruction is the mixing logit ℓ defined in Eq. (62). In contrast to the denoising moments, we do not learn it with a separate neural-network head. Instead, we use an analytic approximation based on a simple parametrization of the two plate distributions.

For $i = 0, 1$, we approximate

$$\mathbb{P}_i(\mathbf{x}) \approx \mathcal{N}(\boldsymbol{\mu}_i, \sigma_i^2 I_D), \qquad \mathbb{P}_0 = \mathbb{P}, \quad \mathbb{P}_1 = \mathbb{Q}, \tag{85}$$

where $\boldsymbol{\mu}_i$ is the mean of the corresponding distribution and σ_i^2 is its mean variance per coordinate, using the same σ_i as in Eq. (77). This approximation is exact for a Gaussian plate and provides a simple moment-matched description for a general data distribution.

Proposition D.6 (Gaussian approximation to the Coulomb weights)**.** *Under the Gaussian approximation* (85)*, the Coulomb weights $\lambda_{i,C}$ defined in Eq.* (60) *satisfy*

$$\lambda_{i,C}(\mathbf{x}, z) \approx \frac{\pi^{\frac{D+1}{2}}}{\Gamma\left(\frac{D+1}{2}\right)} \widetilde{\lambda}_{i,C}(\mathbf{x}, z), \tag{86}$$

where

$$\widetilde{\lambda}_{i,C}(\mathbf{x}, z) = \sum_{n=-C}^{C} \frac{1}{|z - z_{2n+i}|} \mathbb{E}_{u \sim \chi_1^2}\left[\mathcal{N}\left(\mathbf{x}; \boldsymbol{\mu}_i, \left(\sigma_i^2 + \frac{(z - z_{2n+i})^2}{u}\right) I_D\right)\right]. \tag{87}$$

The prefactor in Eq. (86) *is independent of i, $\mathbf{x}$, and z, and therefore cancels in the mixing logit.*

Here χ_1^2 denotes the chi-square distribution with one degree of freedom. The expectation in Eq. (87) is one-dimensional and is evaluated numerically. Proof is given in Appendix E in E.6

Accordingly, throughout the experiments we approximate the mixing logit by substituting Eq. (87) into Eq. (62),

$$\widetilde{\ell}(\mathbf{x}, z) = \log \frac{\widetilde{\lambda}_{0,C}(\mathbf{x}, z)}{\widetilde{\lambda}_{1,C}(\mathbf{x}, z)}. \tag{88}$$

This approximation is exact when both plate distributions are described by Eq. (85), but is only approximate for a general target distribution. Nevertheless, it provides stable mixing weights and performs well in our numerical experiments; we therefore use $\widetilde{\ell}$ rather than a separately learned logit in all reported experiments.

D.5 Full training objective

Combining the three objectives of §5.1,

$$\mathcal{L} = \mathcal{L}_{\mathrm{den}} + \lambda_{\mathrm{ax}}\mathcal{L}_{\mathrm{ax}} + \lambda_{\mathrm{mix}}\mathcal{L}_{\mathrm{mix}}, \tag{89}$$

with $\mathcal{L}_{\mathrm{den}}$, $\mathcal{L}_{\mathrm{ax}}$ and $\mathcal{L}_{\mathrm{mix}}$ as defined there. The main text leaves three quantities free; this appendix fixes them:

- the denoising weights $\kappa_i = c_{\mathrm{out},i}^{-2}$ (Eq. (72)), evaluated through the Gaussian estimate $\widetilde{c}_{\mathrm{out},i}$ of Eq. (75);
- the axial baseline $\widetilde{h}_i$ of Eq. (78), built at the plug-in radius $\widetilde{R}_i$ of Eq. (77) and held constant when differentiating $\mathcal{L}_{\mathrm{ax}}$;
- the mixing logit, for which the experiments use the closed form below rather than the learned residual.

Concretely, all reported runs set $\lambda_{\mathrm{mix}} = 0$ and evaluate the logit analytically as $\widetilde{\ell}$ of Eq. (88), so the network carries four heads rather than five. The splitting given in §5.1 makes the substitution explicit: the geometric term $\log(\rho_0/\rho_1)$ is exact in either case, and only the density ratio $\log(m_0/m_1)$ is replaced, by matching the first two moments of each plate. The approximation is exact for Gaussian plates and introduces an error elsewhere; we keep it because the closed form admits no direction along which a head can drift, which the learned residual does.

D.6 Training procedure

Algorithm 1 C^{2}FM training

Require: sample access to $\mathbb{P}$ and $\mathbb{Q}$; coverage p_z on $(0, L)$; plate gap L; copies C; weights $\lambda_{\mathrm{ax}}, \lambda_{\mathrm{mix}} > 0$; step size α; network f_θ returning $\overline{\mathbf{x}}_{\theta,0}, \overline{\mathbf{x}}_{\theta,1} \in \mathbb{R}^D$ and $g_{\theta,0}, g_{\theta,1}, b_\theta \in \mathbb{R}$

Ensure: transverse moments $\overline{\mathbf{x}}_{\theta,i}$, axial moments $\overline{h}_{\theta,i}$, mixing logit ℓ_θ

repeat
  $\mathbf{x}_0 \sim \mathbb{P}$, $\mathbf{x}_1 \sim \mathbb{Q}$, $z \sim p_z$
  $i \sim \mathrm{Unif}\{0,1\}$, $n \sim \Pr(n \mid i, z) \propto |z - z_{2n+i}|^{-1}$ {side and copy}
  $\mathbf{x} \leftarrow \mathbf{x}_i + |z - z_{2n+i}|\,\boldsymbol{\xi}/|\xi_z|$, $\boldsymbol{\xi} \sim \mathcal{N}(0, I_D)$, $\xi_z \sim \mathcal{N}(0,1)$ {query from the Coulomb kernel}
  $\gamma_0, \gamma_1 \leftarrow$ Eq. (68) {undo the two-sided proposal}
  $(\overline{\mathbf{x}}_{\theta,0}, \overline{\mathbf{x}}_{\theta,1}, g_{\theta,0}, g_{\theta,1}, b_\theta) \leftarrow f_\theta(\mathbf{x}, z)$
  $\mathcal{L}_{\mathrm{den}} \leftarrow$ Eq. (13) {targets $\mathbf{x}_0, \mathbf{x}_1$}
  $\widetilde{h}_i \leftarrow$ Eq. (78), $y_i \leftarrow h_i(\mathbf{x}_i; \mathbf{x}, z)/\widetilde{h}_i$ {$\widetilde{h}_i$ held constant}
  $\mathcal{L}_{\mathrm{ax}} \leftarrow$ Eq. (14)
  $\mathcal{L}_{\mathrm{mix}} \leftarrow$ Eq. (15) {target is the side index $\mathbb{1}[i = 0]$}
  $\theta \leftarrow \theta - \alpha\nabla_\theta\big(\mathcal{L}_{\mathrm{den}} + \lambda_{\mathrm{ax}}\mathcal{L}_{\mathrm{ax}} + \lambda_{\mathrm{mix}}\mathcal{L}_{\mathrm{mix}}\big)$
until converged

E PROOFS OF THE FIELD REPRESENTATION AND LEARNING OBJECTIVES

Theorem E.1 (Denoiser representation of the field)**.** *Let* $a(\mathbf{x}, z) = \sigma(\ell(\mathbf{x}, z))^2$. *Then*

$$\mathbf{E}_C(\mathbf{x}, z) \propto \left(a(\mathbf{x} - \overline{\mathbf{x}}_0) + (1-a)(\overline{\mathbf{x}}_1 - \mathbf{x}),\ a\,\overline{h}_0 + (1-a)\,\overline{h}_1\right). \tag{90}$$

Proof. For convenience, denote the Coulomb kernel associated with the nth copy of side i by

$$K_{i,n}(\mathbf{x}, \mathbf{x}'; z) = \left[\|\mathbf{x} - \mathbf{x}'\|^2 + (z - z_{2n+i})^2\right]^{-\frac{D+1}{2}}. \tag{91}$$

Since $z \in (0, L)$, all kernels are positive and hence $\lambda_{0,C}, \lambda_{1,C} > 0$.

Using Eq. (6), the finite-chain field can be written, up to a common positive prefactor, as

$$\begin{aligned}\mathbf{E}_C(\mathbf{x}, z) \propto \Bigg(&\sum_{n=-C}^{C} \int (\mathbf{x} - \mathbf{x}')\mathbb{P}_0(\mathbf{x}')K_{0,n}\, d^D\mathbf{x}' \\ &+ \sum_{n=-C}^{C} \int (\mathbf{x}' - \mathbf{x})\mathbb{P}_1(\mathbf{x}')K_{1,n}\, d^D\mathbf{x}', \\ &\sum_{n=-C}^{C} \int (z - z_{2n})\mathbb{P}_0(\mathbf{x}')K_{0,n}\, d^D\mathbf{x}' \\ &+ \sum_{n=-C}^{C} \int (z_{2n+1} - z)\mathbb{P}_1(\mathbf{x}')K_{1,n}\, d^D\mathbf{x}' \Bigg).\end{aligned} \tag{92}$$

Now we use the definitions of $\lambda_{i,C}$, $\overline{\mathbf{x}}_i$, and $\overline{h}_i$. For the transverse components,

$$\begin{aligned}\sum_n \int (\mathbf{x} - \mathbf{x}')\mathbb{P}_0 K_{0,n}\, d^D\mathbf{x}' &= \lambda_{0,C}(\mathbf{x} - \overline{\mathbf{x}}_0), \\ \sum_n \int (\mathbf{x}' - \mathbf{x})\mathbb{P}_1 K_{1,n}\, d^D\mathbf{x}' &= \lambda_{1,C}(\overline{\mathbf{x}}_1 - \mathbf{x}).\end{aligned} \tag{93}$$

Similarly, from the definition of $\overline{h}_i$,

$$\begin{aligned}\sum_n \int (z - z_{2n})\mathbb{P}_0 K_{0,n}\, d^D\mathbf{x}' &= \lambda_{0,C}\overline{h}_0, \\ \sum_n \int (z_{2n+1} - z)\mathbb{P}_1 K_{1,n}\, d^D\mathbf{x}' &= \lambda_{1,C}\overline{h}_1.\end{aligned} \tag{94}$$

Therefore,

$$\mathbf{E}_C(\mathbf{x}, z) \propto \left(\lambda_{0,C}(\mathbf{x} - \overline{\mathbf{x}}_0) + \lambda_{1,C}(\overline{\mathbf{x}}_1 - \mathbf{x}),\ \lambda_{0,C}\overline{h}_0 + \lambda_{1,C}\overline{h}_1\right). \tag{95}$$

This expression has a direct electrostatic interpretation: $(\overline{\mathbf{x}}_0, \overline{h}_0)$ determines the effective source of the repulsive contribution from the positive copies, whereas $(\overline{\mathbf{x}}_1, \overline{h}_1)$ determines the effective center towards which the contribution from the negative copies is directed. The factors $\lambda_{0,C}$ and $\lambda_{1,C}$ specify their relative strengths.

It remains only to rewrite these two weights. By Eq. (62),

$$e^{\ell} = \frac{\lambda_{0,C}}{\lambda_{1,C}}, \tag{96}$$

and hence

$$a = \sigma(\ell) = \frac{e^{\ell}}{1 + e^{\ell}} = \frac{\lambda_{0,C}}{\lambda_{0,C} + \lambda_{1,C}}, \qquad 1 - a = \frac{\lambda_{1,C}}{\lambda_{0,C} + \lambda_{1,C}}. \tag{97}$$

[2]where the standard logistic sigmoid $\sigma(u) = (1 + e^{-u})^{-1}$.

Factoring the common positive quantity $\lambda_{0,C} + \lambda_{1,C}$ out of Eq. (95) gives

$$\mathbf{E}_C(\mathbf{x}, z) \propto \left(a(\mathbf{x} - \overline{\mathbf{x}}_0) + (1-a)(\overline{\mathbf{x}}_1 - \mathbf{x}),\ a\,\overline{h}_0 + (1-a)\,\overline{h}_1\right), \tag{98}$$

which proves the statement. □

Proposition E.2 (Coulomb-kernel sampling)**.** *For $i = 0, 1$, Eqs.* (64)–(12) *induce the conditional density*

$$q_i(\mathbf{x} \mid \mathbf{x}_i, z) = \frac{\Gamma\left(\frac{D+1}{2}\right)}{\pi^{\frac{D+1}{2}}} \frac{\sum_{n=-C}^{C} \left(\|\mathbf{x} - \mathbf{x}_i\|^2 + (z - z_{2n+i})^2\right)^{-\frac{D+1}{2}}}{\sum_{n=-C}^{C} |z - z_{2n+i}|^{-1}}. \tag{99}$$

Each copy contributes a multivariate Cauchy kernel, i.e. a Student-t with one degree of freedom, and q_i is their mixture.

Proof. Fix a side i and a copy n, and denote

$$r_n = |z - z_{2n+i}|.$$

For this fixed copy, Eq. (12) gives

$$\mathbf{x} - \mathbf{x}_i = r_n \frac{\boldsymbol{\xi}}{|\xi_z|}, \qquad \boldsymbol{\xi} \sim \mathcal{N}(0, I_D), \quad \xi_z \sim \mathcal{N}(0, 1).$$

Let

$$s = |\xi_z|.$$

Conditioned on s, the random vector $\mathbf{x} - \mathbf{x}_i$ is Gaussian:

$$\mathbf{x} - \mathbf{x}_i \mid s \sim \mathcal{N}\left(0, \frac{r_n^2}{s^2} I_D\right). \tag{100}$$

Hence its conditional density is

$$p(\mathbf{x} \mid \mathbf{x}_i, n, z, s) = \frac{s^D}{(2\pi)^{D/2} r_n^D} \exp\left[-\frac{s^2 \|\mathbf{x} - \mathbf{x}_i\|^2}{2r_n^2}\right]. \tag{101}$$

Since $s = |\xi_z|$ and $\xi_z \sim \mathcal{N}(0, 1)$, s has the half-normal density

$$p(s) = \sqrt{\frac{2}{\pi}} e^{-s^2/2}, \qquad s > 0. \tag{102}$$

Integrating over s, we obtain

$$\begin{aligned} p(\mathbf{x} \mid \mathbf{x}_i, n, z) &= \int_0^\infty p(\mathbf{x} \mid \mathbf{x}_i, n, z, s)\, p(s)\, ds \\ &= \frac{\sqrt{2/\pi}}{(2\pi)^{D/2} r_n^D} \int_0^\infty s^D \exp\left[-\frac{s^2}{2}\left(1 + \frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{r_n^2}\right)\right] ds. \end{aligned} \tag{103}$$

Using

$$\int_0^\infty s^D e^{-As^2/2}\, ds = 2^{\frac{D-1}{2}} \Gamma\left(\frac{D+1}{2}\right) A^{-\frac{D+1}{2}}, \qquad A > 0, \tag{104}$$

we find

$$p(\mathbf{x} \mid \mathbf{x}_i, n, z) = \frac{\Gamma\left(\frac{D+1}{2}\right)}{\pi^{\frac{D+1}{2}}} \frac{r_n}{\left(\|\mathbf{x} - \mathbf{x}_i\|^2 + r_n^2\right)^{\frac{D+1}{2}}}. \tag{105}$$

Thus, a single plate copy generates exactly a multivariate Student distribution with one degree of freedom, i.e. a multivariate Cauchy kernel.

It remains to average over the plate copies. According to Eq. (64),

$$\Pr(n \mid i, z) = \frac{r_n^{-1}}{\sum_{m=-C}^{C} r_m^{-1}}. \tag{106}$$

Therefore,

$$\begin{aligned} q_i(\mathbf{x} \mid \mathbf{x}_i, z) &= \sum_{n=-C}^{C} \Pr(n \mid i, z)\, p(\mathbf{x} \mid \mathbf{x}_i, n, z) \\ &= \frac{\Gamma\left(\frac{D+1}{2}\right)}{\pi^{\frac{D+1}{2}}} \frac{\sum_{n=-C}^{C} \left(\|\mathbf{x} - \mathbf{x}_i\|^2 + r_n^2\right)^{-\frac{D+1}{2}}}{\sum_{m=-C}^{C} r_m^{-1}}. \end{aligned} \tag{107}$$

Substituting back $r_n = |z - z_{2n+i}|$ gives Eq. (66). □

Theorem E.3 (Consistency of the joint denoising objective)**.** *Let $\kappa_i(\mathbf{x}, z) > 0$, $i = 0, 1$, be arbitrary functions independent of the latent samples $\mathbf{x}_i$. Then the population loss*

$$\begin{aligned} \mathcal{L}_{\text{den}} = \mathbb{E}\Big[&\gamma_0 \kappa_0 \left\|\overline{\mathbf{x}}_{\theta,0}(\mathbf{x}, z) - \mathbf{x}_0\right\|^2 \\ &+ \gamma_1 \kappa_1 \left\|\overline{\mathbf{x}}_{\theta,1}(\mathbf{x}, z) - \mathbf{x}_1\right\|^2 \Big] \end{aligned} \tag{108}$$

is minimized pointwise by

$$\overline{\mathbf{x}}_{\theta,0}(\mathbf{x}, z) = \overline{\mathbf{x}}_0(\mathbf{x}, z), \qquad \overline{\mathbf{x}}_{\theta,1}(\mathbf{x}, z) = \overline{\mathbf{x}}_1(\mathbf{x}, z), \tag{109}$$

where the right-hand sides are the Coulomb-weighted means defined in Eq. (61)*.*

Proof. We prove the statement for a fixed side $i \in \{0, 1\}$. The two sides are independent and the same argument applies to both.

For fixed endpoint samples $\mathbf{x}_0, \mathbf{x}_1$ and z, the training query is drawn from the equal mixture

$$q_{\text{mix}}(\mathbf{x} \mid \mathbf{x}_0, \mathbf{x}_1, z) = \frac{1}{2}\left[q_0(\mathbf{x} \mid \mathbf{x}_0, z) + q_1(\mathbf{x} \mid \mathbf{x}_1, z)\right]. \tag{110}$$

By Eq. (68),

$$\gamma_i = \frac{q_i(\mathbf{x} \mid \mathbf{x}_i, z)}{q_{\text{mix}}(\mathbf{x} \mid \mathbf{x}_0, \mathbf{x}_1, z)}. \tag{111}$$

Therefore,

$$q_{\text{mix}}\, \gamma_i = q_i(\mathbf{x} \mid \mathbf{x}_i, z). \tag{112}$$

Thus, the role of γ_i is simply to undo the mixture used to generate the training query.

Let

$$\mathbf{m}_i(\mathbf{x}, z) = \overline{\mathbf{x}}_{\theta,i}(\mathbf{x}, z).$$

Using Eq. (112) and integrating out the endpoint from the opposite side, the contribution of head i to the population loss can be written as

$$\mathcal{L}_i = \mathbb{E}_z \int d^D\mathbf{x}\, \kappa_i(\mathbf{x}, z) \int d^D\mathbf{x}_i\, \mathbb{P}_i(\mathbf{x}_i) q_i(\mathbf{x} \mid \mathbf{x}_i, z) \left\|\mathbf{m}_i(\mathbf{x}, z) - \mathbf{x}_i\right\|^2, \tag{113}$$

where $\mathbb{P}_0 = \mathbb{P}$ and $\mathbb{P}_1 = \mathbb{Q}$.

Since $\kappa_i(\mathbf{x}, z) > 0$ and does not depend on $\mathbf{x}_i$, it cannot change the pointwise minimizer. Hence, for fixed $(\mathbf{x}, z)$ we only need to minimize

$$F_i(\mathbf{m}) = \int d^D\mathbf{x}_i\, \mathbb{P}_i(\mathbf{x}_i) q_i(\mathbf{x} \mid \mathbf{x}_i, z) \|\mathbf{m} - \mathbf{x}_i\|^2. \tag{114}$$

Differentiating with respect to $\mathbf{m}$ gives

$$\nabla_{\mathbf{m}} F_i = 2\mathbf{m} \int \mathbb{P}_i(\mathbf{x}_i) q_i(\mathbf{x} \mid \mathbf{x}_i, z) d^D\mathbf{x}_i - 2 \int \mathbf{x}_i \, \mathbb{P}_i(\mathbf{x}_i) q_i(\mathbf{x} \mid \mathbf{x}_i, z) d^D\mathbf{x}_i. \tag{115}$$

Therefore the unique minimizer is

$$\mathbf{m}_i^*(\mathbf{x}, z) = \frac{\displaystyle\int \mathbf{x}_i \, \mathbb{P}_i(\mathbf{x}_i) q_i(\mathbf{x} \mid \mathbf{x}_i, z) d^D\mathbf{x}_i}{\displaystyle\int \mathbb{P}_i(\mathbf{x}_i) q_i(\mathbf{x} \mid \mathbf{x}_i, z) d^D\mathbf{x}_i}. \tag{116}$$

It remains to identify this ratio. From Proposition E.2,

$$q_i(\mathbf{x} \mid \mathbf{x}_i, z) = A_i(z) \sum_{n=-C}^{C} \left[\|\mathbf{x} - \mathbf{x}_i\|^2 + (z - z_{2n+i})^2 \right]^{-\frac{D+1}{2}}, \tag{117}$$

where

$$A_i(z) = \frac{\Gamma\left(\frac{D+1}{2}\right)}{\pi^{\frac{D+1}{2}}} \left(\sum_{n=-C}^{C} |z - z_{2n+i}|^{-1} \right)^{-1} \tag{118}$$

does not depend on $\mathbf{x}_i$.

The same factor $A_i(z)$ therefore cancels between the numerator and denominator of Eq. (116). We obtain

$$\mathbf{m}_i^*(\mathbf{x}, z) = \frac{\displaystyle\sum_{n=-C}^{C} \int \mathbf{x}_i \, \mathbb{P}_i(\mathbf{x}_i) \left[\|\mathbf{x} - \mathbf{x}_i\|^2 + (z - z_{2n+i})^2 \right]^{-\frac{D+1}{2}} d^D\mathbf{x}_i}{\displaystyle\sum_{n=-C}^{C} \int \mathbb{P}_i(\mathbf{x}_i) \left[\|\mathbf{x} - \mathbf{x}_i\|^2 + (z - z_{2n+i})^2 \right]^{-\frac{D+1}{2}} d^D\mathbf{x}_i}. \tag{119}$$

The denominator is precisely $\lambda_{i,C}(\mathbf{x}, z)$, while the whole ratio is exactly the Coulomb-weighted mean $\overline{\mathbf{x}}_i(\mathbf{x}, z)$ defined in Eq. (61). Hence

$$\overline{\mathbf{x}}_{\theta,i}^* = \overline{\mathbf{x}}_i, \qquad i = 0, 1, \tag{120}$$

which proves the statement. □

Proposition E.4 (Balanced denoising objective)**.** *For every $(\mathbf{x}, z, i)$ such that $c_{\mathrm{out},i}(\mathbf{x}, z) > 0$, the choice* (72) *normalizes the irreducible contribution to the denoising loss to the same scale:*

$$\frac{1}{D} \kappa_i(\mathbf{x}, z) \, \mathbb{E}\Big[\|\mathbf{x}_i - \overline{\mathbf{x}}_i(\mathbf{x}, z)\|^2 \,\Big|\, \mathbf{x}, z, i \Big] = 1. \tag{121}$$

Thus, the network is trained on the displacement relative to its local uncertainty scale rather than being dominated by regions where this displacement is intrinsically large.

Proof. By definition,

$$c_{\mathrm{out},i}^2(\mathbf{x}, z) = \frac{1}{D} \mathbb{E}\Big[\|\mathbf{x}_i - \overline{\mathbf{x}}_i(\mathbf{x}, z)\|^2 \,\Big|\, \mathbf{x}, z, i \Big]. \tag{122}$$

This quantity is the conditional mean squared residual per coordinate after using the optimal transverse denoiser $\overline{\mathbf{x}}_i(\mathbf{x}, z)$.

Using the choice

$$\kappa_i(\mathbf{x}, z) = c_{\mathrm{out},i}^{-2}(\mathbf{x}, z), \tag{123}$$

we immediately obtain

$$\begin{aligned} \frac{1}{D} \kappa_i \mathbb{E}\Big[\|\mathbf{x}_i - \overline{\mathbf{x}}_i\|^2 \,\Big|\, \mathbf{x}, z, i \Big] &= c_{\mathrm{out},i}^{-2} c_{\mathrm{out},i}^2 \\ &= 1. \end{aligned} \tag{124}$$

Hence, at every query point, the irreducible denoising error is measured in units of its own local conditional variance. Regions with a larger intrinsic uncertainty therefore do not receive a larger weight merely because their unavoidable squared error is larger. □

Proposition E.5 (Exact single-sample axial contribution)**.** *For a known source sample $\mathbf{x}_i$ and a query point $(\mathbf{x}, z)$, the exact axial contribution of the corresponding finite chain is*

$$h_i(\mathbf{x}_i; \mathbf{x}, z) = \frac{\displaystyle\sum_{n=-C}^{C} (-1)^i (z - z_{2n+i}) \left[\|\mathbf{x} - \mathbf{x}_i\|^2 + (z - z_{2n+i})^2\right]^{-\frac{D+1}{2}}}{\displaystyle\sum_{n=-C}^{C} \left[\|\mathbf{x} - \mathbf{x}_i\|^2 + (z - z_{2n+i})^2\right]^{-\frac{D+1}{2}}}, \qquad i = 0, 1. \tag{125}$$

Proof. For a fixed source sample $\mathbf{x}_i$, the corresponding plate distribution is a point measure,

$$\mathbb{P}_i(\mathbf{x}') = \delta^{(D)}(\mathbf{x}' - \mathbf{x}_i). \tag{126}$$

Substituting this distribution into the Coulomb weight Eq. (60), we obtain

$$\lambda_{i,C}^{(\mathbf{x}_i)}(\mathbf{x}, z) = \sum_{n=-C}^{C} \left[\|\mathbf{x} - \mathbf{x}_i\|^2 + (z - z_{2n+i})^2\right]^{-\frac{D+1}{2}}. \tag{127}$$

Likewise, substituting the same point measure into the definition of the axial weighted mean in Eq. (61) gives

$$h_i(\mathbf{x}_i; \mathbf{x}, z) = \frac{(-1)^i}{\lambda_{i,C}^{(\mathbf{x}_i)}(\mathbf{x}, z)} \sum_{n=-C}^{C} (z - z_{2n+i}) \left[\|\mathbf{x} - \mathbf{x}_i\|^2 + (z - z_{2n+i})^2\right]^{-\frac{D+1}{2}}. \tag{128}$$

Using Eq. (127) for the denominator immediately gives Eq. (76). □

Proposition E.6 (Gaussian approximation to the Coulomb weights)**.** *Under the Gaussian approximation* (85)*, the Coulomb weights $\lambda_{i,C}$ defined in Eq.* (60) *satisfy*

$$\lambda_{i,C}(\mathbf{x}, z) \approx \frac{\pi^{\frac{D+1}{2}}}{\Gamma\left(\frac{D+1}{2}\right)} \widetilde{\lambda}_{i,C}(\mathbf{x}, z), \tag{129}$$

where

$$\widetilde{\lambda}_{i,C}(\mathbf{x}, z) = \sum_{n=-C}^{C} \frac{1}{|z - z_{2n+i}|} \mathbb{E}_{u \sim \chi_1^2} \left[\mathcal{N}\left(\mathbf{x}; \boldsymbol{\mu}_i, \left(\sigma_i^2 + \frac{(z - z_{2n+i})^2}{u}\right) I_D\right)\right]. \tag{130}$$

The prefactor in Eq. (86) *is independent of i, $\mathbf{x}$, and z, and therefore cancels in the mixing logit.*

Proof. Consider first a single copy of side i and denote

$$r = |z - z_{2n+i}|.$$

Its contribution to the Coulomb weight is

$$\lambda_{i,n}(\mathbf{x}, z) = \int \mathbb{P}_i(\mathbf{x}') \left(\|\mathbf{x} - \mathbf{x}'\|^2 + r^2\right)^{-\frac{D+1}{2}} d^D\mathbf{x}'. \tag{131}$$

We now rewrite the kernel in Eq. (131). From the sampling rule used in Proposition E.2,

$$\mathbf{x} = \mathbf{x}' + r\frac{\boldsymbol{\xi}}{|\xi_z|}, \qquad \boldsymbol{\xi} \sim \mathcal{N}(0, I_D), \quad \xi_z \sim \mathcal{N}(0, 1). \tag{132}$$

Introduce

$$u = \xi_z^2. \tag{133}$$

Since $\xi_z \sim \mathcal{N}(0, 1)$, we have

$$u \sim \chi_1^2. \tag{134}$$

For a fixed value of u, the denominator $|\xi_z| = \sqrt{u}$ is fixed. Therefore

$$\mathbf{x} \mid (\mathbf{x}', u) = \mathbf{x}' + \frac{r}{\sqrt{u}}\boldsymbol{\xi} \tag{135}$$

is Gaussian:

$$\mathbf{x} \mid (\mathbf{x}', u) \sim \mathcal{N}\left(\mathbf{x}', \frac{r^2}{u} I_D\right). \tag{136}$$

Averaging Eq. (136) over $u \sim \chi_1^2$ therefore gives the unconditional density generated by the sampling rule,

$$p_r(\mathbf{x} \mid \mathbf{x}') = \mathbb{E}_{u\sim\chi_1^2}\left[\mathcal{N}\left(\mathbf{x}; \mathbf{x}', \frac{r^2}{u} I_D\right)\right]. \tag{137}$$

On the other hand, Proposition E.2 gives this same density explicitly:

$$p_r(\mathbf{x} \mid \mathbf{x}') = \frac{\Gamma\left(\frac{D+1}{2}\right)}{\pi^{\frac{D+1}{2}}} \frac{r}{\left(\|\mathbf{x} - \mathbf{x}'\|^2 + r^2\right)^{\frac{D+1}{2}}}. \tag{138}$$

Equating Eqs. (137) and (138), we obtain

$$\left(\|\mathbf{x} - \mathbf{x}'\|^2 + r^2\right)^{-\frac{D+1}{2}} = \frac{\pi^{\frac{D+1}{2}}}{\Gamma\left(\frac{D+1}{2}\right)} \frac{1}{r} \mathbb{E}_{u\sim\chi_1^2}\left[\mathcal{N}\left(\mathbf{x}; \mathbf{x}', \frac{r^2}{u} I_D\right)\right]. \tag{139}$$

We now approximate the plate distribution by Eq. (85),

$$\mathbb{P}_i(\mathbf{x}') \approx \mathcal{N}(\mathbf{x}'; \boldsymbol{\mu}_i, \sigma_i^2 I_D). \tag{140}$$

Substituting Eq. (139) into Eq. (131) gives

$$\lambda_{i,n} \approx \frac{\pi^{\frac{D+1}{2}}}{\Gamma\left(\frac{D+1}{2}\right)} \frac{1}{r} \mathbb{E}_{u\sim\chi_1^2}\left[\int \mathcal{N}(\mathbf{x}'; \boldsymbol{\mu}_i, \sigma_i^2 I_D) \mathcal{N}\left(\mathbf{x}; \mathbf{x}', \frac{r^2}{u} I_D\right) d^D\mathbf{x}'\right]. \tag{141}$$

The remaining integral has a simple probabilistic meaning. If

$$\mathbf{x}' \sim \mathcal{N}(\boldsymbol{\mu}_i, \sigma_i^2 I_D) \tag{142}$$

and, conditionally on $\mathbf{x}'$,

$$\mathbf{x} = \mathbf{x}' + \boldsymbol{\eta}, \qquad \boldsymbol{\eta} \sim \mathcal{N}\left(0, \frac{r^2}{u} I_D\right), \tag{143}$$

then the sum of the two independent Gaussian variables is Gaussian with added variances:

$$\mathbf{x} \sim \mathcal{N}\left(\boldsymbol{\mu}_i, \left(\sigma_i^2 + \frac{r^2}{u}\right) I_D\right). \tag{144}$$

Hence

$$\int \mathcal{N}(\mathbf{x}'; \boldsymbol{\mu}_i, \sigma_i^2 I_D) \mathcal{N}\left(\mathbf{x}; \mathbf{x}', \frac{r^2}{u} I_D\right) d^D\mathbf{x}' = \mathcal{N}\left(\mathbf{x}; \boldsymbol{\mu}_i, \left(\sigma_i^2 + \frac{r^2}{u}\right) I_D\right). \tag{145}$$

Therefore

$$\lambda_{i,n} \approx \frac{\pi^{\frac{D+1}{2}}}{\Gamma\left(\frac{D+1}{2}\right)} \frac{1}{r} \mathbb{E}_{u\sim\chi_1^2}\left[\mathcal{N}\left(\mathbf{x}; \boldsymbol{\mu}_i, \left(\sigma_i^2 + \frac{r^2}{u}\right) I_D\right)\right]. \tag{146}$$

Finally, summing over all copies $n = -C, \ldots, C$ and restoring $r = |z - z_{2n+i}|$ gives Eqs. (86)–(87).

The prefactor $\pi^{(D+1)/2}/\Gamma((D+1)/2)$ is the same for $i = 0$ and $i = 1$, so it cancels in the ratio entering the mixing logit. □

F Inference Details

This appendix supplies what §5.2 defers: why the axial clock admits no grid, how the chain field is evaluated, the step rule and its caps, the two axial offsets and the final readout. Algorithm 2 collects the procedure.

F.1 Why the axial clock admits no grid

Fix one atom $\mathbf{x}'$ on side i and look at its own copies. All of them carry the same charge, so each contributes the *same* transverse displacement $\mathbf{x}-\mathbf{x}'$, and the transverse field of the sub-chain is that displacement times the total weight. Their axial displacements $z - z_{2n+i}$, in contrast, come with both signs and cancel. Writing $R = \|\mathbf{x}-\mathbf{x}'\|$, a for the gap to the nearest copy and $\nu = (D+1)/2$ for the kernel exponent, the surviving fraction of the bare axial gap is

$$u(R,a) = \frac{1}{a}\,\frac{\sum_n (a-2nL)\big[R^2+(a-2nL)^2\big]^{-\nu}}{\sum_n \big[R^2+(a-2nL)^2\big]^{-\nu}} \in (0,1], \tag{147}$$

and the right-hand side of (17) is inflated by exactly $1/u$. This is the confinement of Theorem 3.1 seen from the other side: $u>0$ always, but nothing bounds it from below.

Written as Fourier modes in z rather than as a sum over copies, the same series has u decaying as $\exp\!\big(-\pi^2 R^2/(4\nu L^2)\big)$ up to an algebraic prefactor, so the cancellation becomes severe once

$$R \gtrsim \frac{2L}{\pi}\sqrt{\nu}, \tag{148}$$

which is $R/L \approx 12$ at $D=768$ and $R/L \approx 25$ at $D=3072$. Note the direction: the screening scale grows as $\sqrt{D}$, so the difficulty is *worse* in lower dimension, not better. Figure 6 evaluates (147) exactly.

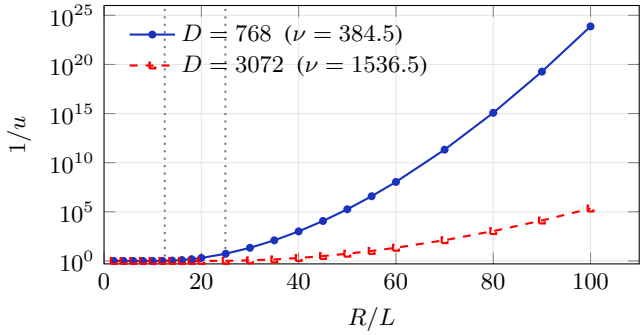


Figure 6: The factor $1/u$ by which the chain inflates the right-hand side of the axial clock (17), from an exact evaluation of (147) at $a = L/2$, $L = 20$, in 400-digit arithmetic. Dotted lines mark the screening scale (148). Below it the axial push is intact; above it a step in z must shrink by the plotted factor.

Along a trajectory the quantity that matters is $u_z = E_z/\|\mathbf{E}\|$, the axial component of the unit direction: a step h of arc advances z by $h\,u_z$, so $1/u_z$ is the price of a unit of z at that point. A grid in z fixes that price in advance, whereas R, and with it $1/u_z$, changes as the line travels, and at the upper end of the range in Figure 6 the axial component is destroyed by cancellation in double precision, where (17) has no reliable value at all. The same holds for the exact field, so this is a property of the construction and not of the network. Arc length avoids the issue because its step rule reads the geometry off the field at the current point instead of committing to it beforehand.

F.2 Evaluating the chain field

At inference the objects of Theorem 5.1 are assembled from the network outputs, and the only sum taken explicitly is the closed-form part of the logit, $\rho_i(z) = \sum_{n=-C}^{C} |z - z_{2n+i}|^{-1}$, together with the axial baseline $\widetilde{h}_i$ of (78). Both are copy sums, and the remark that makes them cheap is that

truncation is never the binding error: the weight of copy n falls as $\exp(-4\nu n^2 L^2/R^2)$, so a handful of copies carries the sum at any R, and $C = 32$ is exact to double precision far beyond the range of (148).

What does fail is cancellation. Once u is of the order of machine epsilon, no subtraction of positive terms in double precision can produce it, and the copy sum returns noise rather than a small number, a failure of the *representation* and not of the truncation. The Fourier form of the same infinite sum has the screened value as its leading term, so it has nothing to cancel, and the image runs evaluate it there instead, switching on R/L (at $R/L = 50$, where the two representations agree), with the Bessel functions of order $\nu - \frac{1}{2}$ taken from the uniform asymptotic expansion, whose cost does not grow with ν.

The switch is what the image experiments need and the two-dimensional ones do not, though not because the chain screens less there. At $D = 2$ the exponent is $\nu = 3/2$ and the scale (148) sits at $R/L \approx 0.8$: evaluating (147) exactly gives $u = 0.37$ already at $R/L = 1$ and $1.5 \cdot 10^{-5}$ at $R/L = 5$, so the axial clock is hopeless there too, and by a wider margin. What differs is the arithmetic. A slowly falling kernel spreads the weight over many copies, so a truncated sum is limited by what it omits rather than by what it cancels; the two-dimensional runs take the $2C + 1$ copies literally, and this is also what makes C a meaningful axis to sweep in those figures.

F.3 Step size, integrator and stopping

All experiments use $\eta = 0.1$ in (19). Two caps act on the resulting h: a step may move z by at most half the distance to the nearer plate, which is what makes the discrete flow inherit Theorem 3.1; and the last step of a trajectory is shortened so that it terminates exactly at z_{stop}. Each step is Heun's method, which costs two evaluations of f_θ, and falls back to its Euler predictor whenever the second stage returns a non-finite direction, so that a single bad probe costs accuracy on one step rather than the trajectory.

The step count is therefore not fixed in advance but is whatever the local geometry produces, subject to a cap of 1000 steps. A line whose norm exceeds $10^{14} L$ is declared escaped and reported as such rather than clipped, since clipping would hide exactly the failure mode the chain is meant to remove.

F.4 The two axial offsets

The offsets z_{start} and $L - z_{\mathrm{stop}}$ enter the method quite differently.

z_{stop} **is a tolerance.** The raw endpoint of a trajectory misses its target atom by an amount proportional to $L - z_{\mathrm{stop}}$, with a constant of proportionality that itself grows with L. An offset specified as a fraction of L therefore degrades as the gap widens, and we specify it absolutely: $L - z_{\mathrm{stop}} = 10^{-3}$ in every experiment, independent of L.

z_{start} **has a window, from both sides.** Below, the network must resolve the field at the source plate, where the Coulomb kernel is at its stiffest; too small an offset asks for accuracy the trained heads do not have. Above, the launch must still identify the source atom it came from: the Coulomb posterior at $(\mathbf{x}, z)$ concentrates on the launching atom for small z and spreads as z grows, and the pairing between a generated sample and its source degrades with it. Every experiment takes $z_{\mathrm{start}} = 0.02$, which is also where the training coverage of the source plate begins.

F.5 Final readout

A trajectory stops at $z_{\mathrm{stop}} < L$, so its raw endpoint still carries the residual described above. Instead of returning it we return the network's own target moment at that point, $\overline{\mathbf{x}}_{\theta,1}(\mathbf{x}, z_{\mathrm{stop}})$. By Theorem 5.1 this is the Coulomb-weighted conditional mean of the target side given the query, and as $z \to L$ that weight concentrates on the atoms closest to the endpoint, so the readout is the posterior mean of the target the line has arrived at. It plays the role of the final denoising step of a diffusion sampler, costs no extra evaluation, since the moment is already computed for the last direction, and is what every experiment reports.

F.6 Inference procedure

Algorithm 2 C^{2}FM inference

Require: trained network f_θ; plate gap L; copies C; step fraction η; offsets $0 < z_{\text{start}} < z_{\text{stop}} < L$
Ensure: a sample approximating $\mathbb{Q}$, paired with its source

$\mathbf{x}_0 \sim \mathbb{P}$, $\boldsymbol{\xi} \sim \mathcal{N}(0, I_D)$, $\xi_z \sim \mathcal{N}(0,1)$
$z \leftarrow z_{\text{start}}$, $\mathbf{x} \leftarrow \mathbf{x}_0 + z_{\text{start}}\, \boldsymbol{\xi}/|\xi_z|$ {flux-weighted crossing law}
while $z < z_{\text{stop}}$ **do**
  $\mathbf{u} \leftarrow \mathbf{E}_\theta(\mathbf{x}, z)/\|\mathbf{E}_\theta(\mathbf{x}, z)\|$ {one pass of f_θ, then Eq. (11)}
  $d_i \leftarrow \sqrt{\text{gap}_i^2 + \|\mathbf{x} - \overline{\mathbf{x}}_{\theta,i}\|^2}$, $h \leftarrow \eta \min(d_0, d_1)$ {Eq. (19); the means are already computed}
  $h \leftarrow \min\{h,\ \frac{1}{2}\min(z, L-z)/|u_z|\}$ {a step never crosses a plate}
  $\mathbf{u}' \leftarrow \mathbf{E}_\theta/\|\mathbf{E}_\theta\|$ at $(\mathbf{x}, z) + h\,\mathbf{u}$
  $(\mathbf{x}, z) \leftarrow (\mathbf{x}, z) + \frac{h}{2}(\mathbf{u} + \mathbf{u}')$ {Heun; the Euler predictor if $\mathbf{u}'$ is not finite}
end while
shorten the last step linearly so that $z = z_{\text{stop}}$
return $\overline{\mathbf{x}}_{\theta,1}(\mathbf{x}, z_{\text{stop}})$ {final readout, §F.5}

G Technical Details

The code reproducing every experiment is included in the supplementary material.

G.1 Data

The two-dimensional datasets are generated procedurally. The 5-GMM $\to$ 4-GMM task uses source means $(-8,8), (8,8), (0,0), (-8,-8), (8,-8)$ and target means $(-8,6), (8,6), (-8,-6), (8,-6)$, with isotropic standard deviation 0.8 for all components. The 1-GMM $\to$ 2-GMM task uses source mean $(0,0)$ with standard deviation 2.0, and target means $(-6,0), (6,0)$ with standard deviation 1.0. Mixture weights are uniform. Both are embedded in $\mathbb{R}^3$ by the auxiliary coordinate, with the source at $z = 0$ and the target at $z = L$.

The image tasks use colored MNIST at 16×16, obtained from the MNIST training split by bilinear resizing. Colour is applied to a grayscale digit at load time by sampling a hue, so the target measure is the distribution over digits *and* colours; the hues are drawn from the global NumPy generator, which we seed explicitly, since two processes that do not agree on it are training against different measures. Two tasks are reported: noise $\to$ colored digits, over all $60\,000$ training images, and the translation $2 \to 3$, whose source is the $5\,958$ images of digit 2 and whose target is the $6\,131$ images of digit 3.

We report FID against a reference drawn from the same training split that the model is trained on; no held-out split is used. Memorisation is therefore not visible in the reported numbers, and they should be read as a measure of how well the transport matches the target measure, not as evidence of generalisation.

G.2 Architecture and heads

The backbone is the EDM UNet used by EFM, with 128 base channels, channel multipliers $(2, 2, 2)$, four residual blocks per resolution, attention at the 16×16 resolution and dropout 0.13: 59.8M parameters. Its input is preconditioned in the EDM manner with the data scale σ_{data} measured on the target archive.

Theorem 5.1 asks for five objects, and the network emits them from one trunk: the output branch carries $2D$ channels for the two transverse moments $\overline{\mathbf{x}}_{\theta,0}, \overline{\mathbf{x}}_{\theta,1}$ and two scalars $g_{\theta,0}, g_{\theta,1}$ for the axial corrections. Since all reported runs take the mixing logit in closed form ($\lambda_{\text{mix}} = 0$), no logit head is instantiated at all, and the scalar b_θ of Algorithm 1 is absent from the trained networks. The transverse moments are not predicted directly but through the posterior skip-and-scale parameterisation of §D, without which the objective is unbounded at the source plate; the axial heads are the exponential corrections $\overline{h}_{\theta,i} = \widetilde{h}_i e^{g_{\theta,i}}$ over the closed-form baseline.

Every sum over the $2C + 1$ copies is evaluated in double precision regardless of the precision the network runs at: the axial baseline, ρ_i and the training proposal alike. They are differences of positive terms whose cancellation is the very quantity being computed (§F.2), and in single precision they return noise where the answer is small.

G.3 Optimisation

The two-dimensional experiments and the validation figures of §3.3 use the *exact* field: $\mathbf{E}_C$ is assembled directly from the empirical source and target measures, with no network anywhere in the loop. They isolate the construction from the learning, and nothing in this subsection applies to them. The paragraphs below describe the two trained image runs.

All trained runs use Adam at learning rate 2×10^{-4} with a linear warmup over the first 1000 iterations, no weight decay, gradient clipping at 10, an exponential moving average of the weights with decay 0.999, and a global batch of 128; under data parallelism the per-rank batch is set so that the global batch is 128 regardless of the number of devices. The loss weights are $\lambda_{\mathrm{ax}} = 1$ and $\lambda_{\mathrm{mix}} = 0$. Training is capped at $200\,000$ iterations; on two A100-80GB cards that is about 17 hours per run including periodic evaluation, measured on a shared machine and therefore an upper bound. Table 1 collects the per-experiment settings.

Experiment	field	D	L	C	best ckpt
1 GMM $\to$ 2 GMM §3.3	exact	2	20	$2, 4, 20$	—
5 GMM $\to$ 4 GMM §3.3	exact	2	20	$0, 2, 4, 20$	—
colored MNIST, both tasks §3.3	exact	768	20	$0, 32$	—
noise $\to$ colored MNIST §5.3	learned	768	20	32	200 000 it.
colored MNIST $2 \to 3$ §5.3	learned	768	20	32	70 000 it.

Table 1: Per-experiment settings. D is the data dimension, L the plate gap, C the number of copies on each side of the chain; *field* says whether $\mathbf{E}_C$ was assembled from the empirical measures or from a trained network. The two-dimensional runs take the chain as the literal sum over the $2C + 1$ copies; the image runs additionally switch to the Fourier form of the same sum past $R/L = 50$ (§F.2). Everything not listed here is shared across experiments and given in the text above.

The Flow Matching baseline of §5.3 is trained under exactly these settings: the same data, the same backbone, the same optimiser, schedule, batch and iteration budget. It differs in the output branch alone, which carries a single D-dimensional velocity in place of the two transverse moments and the two axial scalars, and it received no separate tuning of its own.

G.4 Evaluation

FID is computed on $10\,000$ generated samples with the pool3 features of InceptionV3 (`pytorch-fid` 0.3.0) and the usual Fréchet formula, against reference statistics taken over the training archive.

The FID of a run oscillates from checkpoint to checkpoint by more than the differences we are measuring, so we do not report the last checkpoint. For each run we evaluate the checkpoints saved over training and report the best one, together with the iteration it came from. The sampler settings used for every number are given in §F: the arc-length clock, $\eta = 0.1$, the launch law, the two offsets and the final readout.