How Learning Dynamics Shape Decoding in Masked Diffusion Models

Anonymous authors
Paper under double-blind review

Abstract

Diffusion Language Models have recently emerged as a promising alternative for reasoning, benefiting from bidirectional prediction and flexible token ordering at inference. However, how training shapes the usefulness of this flexibility remains unclear. In this paper, we study stage-wise learning dynamics in a masked diffusion model using a two-layer Transformer with a fixed output head on a graph-planning task. A local endpoint-first asymmetry under ordinary joint training motivates a framework separating *endpoint preparation* from *predecessor learning*. Within this framework, we establish a finite predecessor-risk decrease and derive risk-dependent reverse-circuit certificates. We derive conditions for endpoint reveal to bootstrap predecessor prediction and for top-probability and top-margin decoding to favor the reverse order, distinguishing sequence-level correctness from confidence-based selection. We also characterize order independence and confidence agreement when all mask-conditioned risks vanish. Together, our analysis relates stage-wise learning and internal computation to decoding preference and advantage.

1 Introduction

Despite substantial progress in reasoning, large language models can struggle when left-to-right generation conflicts with a problem's structure (Saparov & He, 2023). In graph planning, choosing an initial move may require identifying which branch reaches the destination, whereas tracing backward can reduce it to local edge retrieval (Trainin et al., 2026). More broadly, discrete diffusion models outperform autoregressive counterparts on graph planning, Countdown, Sudoku, and Boolean satisfiability (Ye et al., 2025), and masked diffusion can mitigate the reversal curse (Berglund et al., 2024) observed in much larger autoregressive models (Nie et al., 2025a). These results motivate interest in masked diffusion models (MDMs), which recover missing tokens from bidirectional context and permit flexible generation orders (Zheng et al., 2025).

Yet flexibility alone does not determine an effective reasoning procedure: the model must choose which token to generate next. Because this choice changes the context for subsequent predictions, decoding becomes part of reasoning itself. Adaptive ordering can avoid difficult masking problems (Kim et al., 2025), but generation quality and efficiency depend on the unmasking policy and sampling schedule (Hayakawa et al., 2026; Li et al., 2026).

Accordingly, existing work develops learned unmasking policies (Hong et al., 2026), analyzes accuracy–parallelism trade-offs (Chen et al., 2026a), and adapts how many tokens to reveal at each step (Ben-Hamu et al., 2025). Other approaches align training with inference-time trajectories (Peng et al., 2026). These approaches address training and decoding design, but leave open how learning creates the conditional prediction abilities and confidence differences that make an order useful. The question is therefore not only which order works, but how it becomes effective through training.

A learning-dynamics perspective offers a way to address this gap. Existing research studies how optimization shapes learned features (Zhang et al., 2025; Chen et al., 2026b; Bordelon et al., 2024; Pesme & Flammarion, 2023) and computational circuits (Olsson et al., 2022; Nichani et al., 2024). For MDMs, this suggests examining how the model learns to exploit revealed tokens: an order can help only insofar as the acquired computation makes use of the resulting context. To address this gap, we ask:

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.

How do learning dynamics shape which decoding orders are effective, and under what conditions does their choice matter?

To explain a planning advantage, the analysis must retain a task that challenges left-to-right prediction. To this end, we study a controlled two-token version of path-star graph planning (Bachmann & Nagarajan, 2024), a task family where next-token-trained models can fail while diffusion models succeed through reverse generation (Trainin et al., 2026). Given a directed graph $G(R, L)$ and a source–destination query, a two-layer Transformer predicts the predecessor and a copy of the destination. Masking either or both answer positions yields four conditional prediction problems.

To preserve the computation behind this planning asymmetry, our reduction retains two-layer softmax attention, as in related planning analyses (Huang et al., 2026): edge encoding must be composed with destination-conditioned retrieval, rather than replaced by a single token association. Retaining this composition, however, couples learning across layers; shared parameters further couple mask-conditioned predictions, while decoding changes the state encountered next. These interactions make a direct joint convergence analysis difficult.

Rather than solve the full joint gradient flow, we characterize what attained low risk implies about reverse-circuit structure and its use in decoding. To study how this computation develops, we use a local endpoint-first loss asymmetry under joint training to motivate stage-wise gradient flow: endpoint preparation followed by predecessor learning. This isolates the learning mechanism without claiming that staging guarantees the required low-risk regimes. Experiments then examine whether these circuit and decoding patterns also emerge under ordinary joint training, beyond the prescribed stages. Our main contributions are as follows.

- **From attained risk to reverse computation.** Task symmetries and stage-wise analysis yield risk-dependent certificates for a reverse-retrieval circuit that matches the destination and retrieves its predecessor.
- **Endpoint reveal and reverse-order selection.** We characterize how revealing a correct endpoint can bootstrap predecessor retrieval through an explicit destination cue, and give sufficient conditions for top-probability and top-margin decoding to select this order.
- **Accuracy advantage and order independence.** We give conditions under which endpoint-reveal bootstrap yields a reverse-order accuracy advantage, distinct from confidence preference, and characterize order independence and confidence agreement as all mask-conditioned risks vanish.

1.1 Related Work

Reverse reasoning circuits and reverse decoding. Bachmann & Nagarajan (2024) use graph planning to expose difficulties of next-token prediction and the benefits of reversing the prediction order. Trainin et al. (2026) empirically show that discrete diffusion models exploit planning asymmetries through reverse decoding, while Huang et al. (2026) analyze how multi-token prediction induces reverse reasoning through a decoupled training signal. These notions of reversal are related but distinct: reverse association, backward retrieval inside a network, and reverse output generation need not coincide. Our work characterizes internal reverse retrieval and separately states the conditions under which confidence-based decoding selects and benefits from reverse generation.

Masked diffusion and decoding policies. Discrete diffusion (Austin et al., 2021; Lou et al., 2024) and masked formulations (Sahoo et al., 2024; Shi et al., 2024) support alternatives to left-to-right generation, including large-scale language modeling (Nie et al., 2025b). Kim et al. (2025) identify the trade-off between training on difficult masking problems and avoiding them during adaptive inference. Subsequent work develops unmasking planners (Peng et al., 2025), information-theoretic analyses of order sensitivity and parallelization (Zhang et al., 2026), and learned scheduling policies (Xu et al., 2026). Kim et al. (2026) instead modify the forward masking process to align training with inference. We study a complementary question: how mask-conditioned learning and attention structure relate to the prediction and confidence properties used by these decoding procedures.

Learning dynamics and finite-time behavior. Analyzing parameter updates provides a way to explain behavior that cannot be inferred from representational capacity or limiting solutions

alone (Ren & Sutherland, 2025). Gradient-based theories characterize the emergence of causal and induction circuits (Chen et al., 2024; Bietti et al., 2023), the selection between retrieval mechanisms and positional shortcuts (Kawata et al., 2025), and how intermediate supervision enables reasoning computations (Kim & Suzuki, 2025; Huang et al., 2025). Several works further emphasize finite-time behavior, including early feature alignment (Min et al., 2024), learning dynamics under finite stopping criteria (Yeom et al., 2026), and stage-wise behavior in diffusion training (Bonnaire et al., 2025; Wang & Pehlevan, 2025). For MDMs, Huang & Mirzasoleiman (2026) analyze how masking regimes shape feature learning and implicit regularization. Our work connects this learning-dynamics perspective to inference by relating intermediate circuit formation to decoding-order selection and correctness.

2 Problem Formulation

2.1 Notation

For a positive integer n, write $[n] = \{1, \dots, n\}$. We use $\mathbf{e}_i$ for standard basis vectors, with dimension determined by context, $\mathbf{I}_n$ for the identity matrix, and $\|\cdot\|_F$ for the Frobenius norm. Softmax is applied row-wise. For a sequence $\mathbf{x}_0$ and a position set M, $\mathbf{x}_0[M]$ replaces exactly the positions in M by [MASK].

2.2 Two-layer Masked Diffusion Model

We use two bidirectional single-head attention layers to isolate how retrieval and transport contribute to prediction. For a masked sequence $\widetilde{\mathbf{x}} = \mathbf{x}_0[M]$, define the fixed one-hot token and positional representation matrices $\mathbf{Z} = [\mathbf{e}_{\widetilde{x}_1}, \dots, \mathbf{e}_{\widetilde{x}_T}]^\top \in \mathbb{R}^{T\times D}$ and $\mathbf{P} = [\mathbf{e}_1, \dots, \mathbf{e}_T]^\top = \mathbf{I}_T \in \mathbb{R}^{T\times T}$, where $D = K + 1$ includes the mask token.

Each layer has disentangled block-diagonal weights $\mathbf{W}^{(\ell)} = \mathrm{diag}(\mathbf{W}_c^{(\ell)}, \mathbf{W}_p^{(\ell)})$, with $\mathbf{W}_c^{(\ell)} \in \mathbb{R}^{D\times D}$ and $\mathbf{W}_p^{(\ell)} \in \mathbb{R}^{T\times T}$. These four attention blocks form the trainable parameters θ. The shared output head is fixed at $\mathbf{U} = [\alpha \mathbf{I}_K \ \mathbf{0}] \in \mathbb{R}^{K\times D}$, with $\alpha > 0$, so improvements in prediction must come from the attention layers. The forward pass is

$$\mathbf{S}^{(1)} = \mathrm{softmax}\Big(\mathbf{Z}\mathbf{W}_c^{(1)}\mathbf{Z}^\top + \mathbf{W}_p^{(1)}\Big), \qquad \mathbf{S}^{(2)} = \mathrm{softmax}\Big(\mathbf{S}^{(1)}\Big(\mathbf{Z}\mathbf{W}_c^{(2)}\mathbf{Z}^\top + \mathbf{W}_p^{(2)}\Big)\Big),$$
$$\mathbf{\Pi} = \mathrm{softmax}\Big(\mathbf{S}^{(2)}\mathbf{S}^{(1)}\mathbf{Z}\mathbf{U}^\top\Big). \tag{1}$$

The i-th row of $\mathbf{\Pi}$ gives the predicted token distribution at position i. To describe information transport, write $\mathbf{H}^{(1)} = \mathbf{S}^{(1)}\mathbf{Z}, \mathbf{Q} = \mathbf{S}^{(2)}\mathbf{S}^{(1)}, \mathbf{H}^{(2)} = \mathbf{Q}\mathbf{Z}$. Here $\mathbf{H}^{(1)}$ contains the Layer-1 token values, $\mathbf{Q}$ the two-layer transport, and $\mathbf{H}^{(2)}$ the values before the output head. An entry $H_{i,v}^{(\ell)}$ is token mass, not the final softmax probability.

Training objective. Rather than fixing a left-to-right prefix, training varies the visible answer context. Draw $k \sim \mathrm{Unif}\{1, 2\}$ and $M \mid k \sim \mathrm{Unif}\{M \subseteq \{a_1, a_2\} : |M| = k\}$, and minimize

$$\mathcal{L}(\theta) = \mathbb{E}_{\mathbf{x}_0\sim\mathrm{Unif}(\mathcal{D})}\mathbb{E}_M\left[\frac{1}{|M|}\sum_{i\in M} -\log p_\theta(x_{0,i} \mid \mathbf{x}_0[M])\right], \tag{2}$$

where $p_\theta(x_{0,i} \mid \mathbf{x}_0[M])$ is the probability assigned to the correct token by output row i. Thus each answer is trained with and without the other revealed: the same parameters must solve all four mask-conditioned problems.

2.3 Minimal Star-Graph Task

Fix integers $R \geq 2$ and $L \geq 3$. The graph $G(R, L)$ consists of R directed arms of $L - 1$ edges sharing a root u. Write $v_{h,0} = u$ for $h \in [R]$, with all non-root vertices $v_{h,j}$ distinct. The edge set is $\mathcal{E} = \{(v_{h,j-1}, v_{h,j}) : h \in [R],\ j \in [L-1]\}$, so $m = |\mathcal{E}| = R(L-1)$ and the number of vertices is $K = 1 + R(L-1)$.

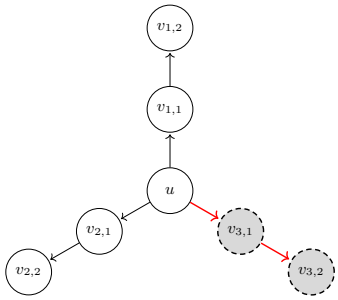
Let $\mathcal{V}$ be a node-token vocabulary of size K, and identify each structural vertex with its assigned token. For a target arm $h_\star \in [R]$, write $v_j = v_{h_\star,j}$. Let $\pi(\mathcal{E})$ denote the concatenation of all directed

edge pairs in a chosen serialization order. The prompt specifies the graph, root, and destination, while the answer lists the destination's immediate predecessor followed by the destination:

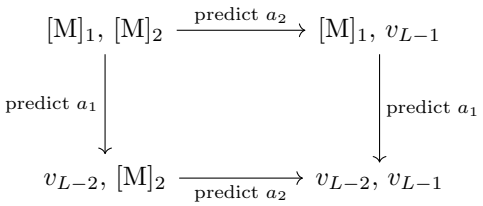
$$\mathbf{x}_0 = \big[\pi(\mathcal{E}) \mid u, v_{L-1} \mid v_{L-2}, v_{L-1}\big], \qquad \mathbf{x}_{\text{inf}} = \big[\pi(\mathcal{E}) \mid u, v_{L-1} \mid [\texttt{MASK}]^2\big]. \tag{3}$$

We sample uniformly from the population $\mathcal{D}$ of all unique clean sequences obtained from all bijective vertex assignments to $\mathcal{V}$, all R target-arm choices, and all directed edge serializations. Only the two answer positions are corrupted, leaving the graph and query visible; Figure 1 shows the resulting reveal orders.[1]

Terminal Edge on the Target Arm. The sequence length is $T = 2m + 4 = 2R(L-1) + 4$. For each $r \in [m]$, let $s_r, d_r \in [T]$ denote the source and destination position of the r-th serialized edge. The query and answer positions are (q_s, q_t) and (a_1, a_2), respectively. We group positions as $\mathcal{I} = (\mathcal{I}_1, \ldots, \mathcal{I}_m, \mathcal{I}_{\text{sp}})$, where $\mathcal{I}_r = (s_r, d_r)$ and $\mathcal{I}_{\text{sp}} = (q_s, q_t, a_1, a_2)$. The unique terminal edge $r_\star \in [m]$ on the target arm satisfies $[\mathbf{x}_0]_{d_{r_\star}} = [\mathbf{x}_0]_{q_t} = [\mathbf{x}_0]_{a_2} = v_{L-1}$ and $[\mathbf{x}_0]_{s_{r_\star}} = [\mathbf{x}_0]_{a_1} = v_{L-2}$.



(a) Example of $G(3,3)$.



(b) Four mask-conditioned prediction states.

Figure 1: Terminal-pair task overview. (a) The red directed path denotes the target arm in $G(3,3)$, and the gray dashed nodes indicate the answer positions. (b) shows the two decoding paths: Left-to-Right(LTR) reveal a_1 first, and Right-to-Left(RTL) reveal a_2 first.

2.4 Symmetry Reduction

Relabeling vertices or reordering serialized edge pairs preserves the task, so population gradient flow, $\dot{\theta}(t) = -\nabla\mathcal{L}(\theta(t))$, preserves these symmetries from $\theta(0) = 0$. Consequently, content weights distinguish equality, while positional weights distinguish source/destination roles and same/different edges. The population orbit and invariance are proved in Proposition A.3 and Lemma A.4.

The flow further preserves Layer-2 pair blindness Lemma A.5: positional scores cannot distinguish individual edges within a graph-role pool, so content must distinguish the relevant destination.

Exact fixed weight structure. Use $\mathbf{W}[\mathcal{J}, \mathcal{K}]$ to select rows $\mathcal{J}$ and columns $\mathcal{K}$. Recall $\mathcal{I}_r = (s_r, d_r)$ and $\mathcal{I}_{\text{sp}} = (q_s, q_t, a_1, a_2)$.

Lemma 2.1 (Exact fixed spaces and softmax-identifiable quotients)**.** *Every content and positional block has the following form for all $t \geq 0$:*

$$\mathbf{W}_c^{(\ell)} = \begin{pmatrix} \alpha_\ell \mathbf{I}_K + \beta_\ell(\mathbf{1}_K\mathbf{1}_K^\top - \mathbf{I}_K) & \rho_\ell \mathbf{1}_K \\ \varsigma_\ell \mathbf{1}_K^\top & \mu_\ell \end{pmatrix}, \tag{4}$$

$$\mathbf{W}_p^{(\ell)} = \begin{pmatrix} \mathbf{I}_m \otimes \mathbf{A}_=^{(\ell)} + (\mathbf{1}_m\mathbf{1}_m^\top - \mathbf{I}_m) \otimes \mathbf{A}_\neq^{(\ell)} & \mathbf{1}_m \otimes \mathbf{B}^{(\ell)} \\ \mathbf{1}_m^\top \otimes \mathbf{C}^{(\ell)} & \mathbf{D}^{(\ell)} \end{pmatrix}. \tag{5}$$

The same-edge block is $\mathbf{A}_=^{(\ell)} = \mathbf{W}_p^{(\ell)}[\mathcal{I}_r, \mathcal{I}_r]$, and the different-edge block is $\mathbf{A}_\neq^{(\ell)} = \mathbf{W}_p^{(\ell)}[\mathcal{I}_r, \mathcal{I}_{r'}]$ for $r \neq r'$. The remaining blocks are $\mathbf{B}^{(\ell)} = \mathbf{W}_p^{(\ell)}[\mathcal{I}_r, \mathcal{I}_{\text{sp}}]$, $\mathbf{C}^{(\ell)} = \mathbf{W}_p^{(\ell)}[\mathcal{I}_{\text{sp}}, \mathcal{I}_r]$, and $\mathbf{D}^{(\ell)} = \mathbf{W}_p^{(\ell)}[\mathcal{I}_{\text{sp}}, \mathcal{I}_{\text{sp}}]$, independently of the choices of r and $r' \neq r$.

[1] Masking additional path positions introduces longer multi-step retrieval and composition chains, which can require deeper computation and substantially complicate an exact analysis of the learning dynamics.

Symmetry thus leaves a graph-size-independent number of coordinates without selecting the dominant interactions. We next connect these interactions to learning and attained risk.

3 Stage-Wise Learning of a Reverse-Reasoning Circuit

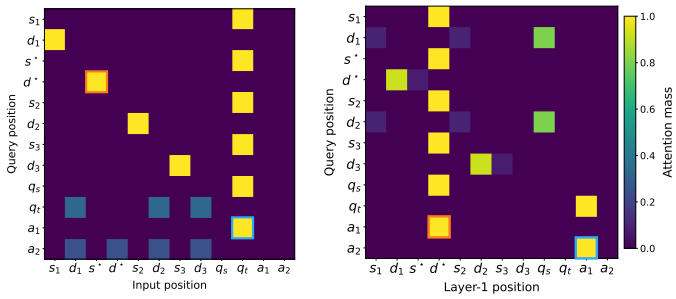
Within the symmetry-reduced model, we analyze a *reverse-reasoning circuit* through two prescribed stages. We first describe both routes on the both-masked input $\widetilde{\mathbf{x}} = \mathbf{x}_0[\{a_1, a_2\}]$ and write $s^\star = s_{r_\star}$, $d^\star = d_{r_\star}$ for the source and destination positions of the target terminal edge.

Stage I (endpoint preparation) supplies endpoint copying from q_t to a_2 through Layer-1 carriers, namely positions that place high first-layer attention on q_t:

$$\mathcal{A}_\varepsilon(t; \widetilde{\mathbf{x}}) := \left\{ k \in [T] : S^{(1)}_{k,q_t}(t; \widetilde{\mathbf{x}}) \geq 1 - \varepsilon \right\}, \qquad 0 < \varepsilon < 1. \tag{6}$$

Stage II (predecessor learning) retains this computation while learning the reverse route: Layer 2 at a_1 matches $d^\star$, whose Layer-1 value comes from its paired source $s^\star$.

With $\mathcal{A} = \mathcal{A}_\varepsilon(t; \widetilde{\mathbf{x}})$, Figure 2 places these stage-wise routes alongside attention maps from joint training. Arrows trace attention from output to input; *reverse* refers to destination-to-source retrieval.



(a) Layer-1 attention (b) Layer-2 attention (c) Stage-wise routes

Figure 2: Reverse-reasoning circuit on the both-masked input. (a–b) Population-averaged attention after joint training. (c) Stage-wise endpoint and predecessor routes.

3.1 Endpoint-First Asymmetry under Joint Training

For $i \neq j$, let $\mathcal{L}_{i|j}(t)$ denote the population loss for predicting a_i with the correct a_j revealed, and let $\mathcal{L}_{i|\varnothing}(t)$ denote the loss with both answers masked. With $\mathcal{E}_4 = (1 \mid \varnothing, 1 \mid 2, 2 \mid \varnothing, 2 \mid 1)$, the objective is

$$\mathcal{L}(\theta(t)) = \frac{1}{4} \sum_{e \in \mathcal{E}_4} \mathcal{L}_e(t). \tag{7}$$

Shared parameters allow cross-task help or interference (Appendix B.1). To address this, we first ask which prediction the joint flow favors.

The loss floor is $\ell_{\min} = \log(1 + (K-1)e^{-\alpha})$. Write $R_e = \mathcal{L}_e - \ell_{\min}$ and

$$R_E = R_{2|\varnothing} + R_{2|1}, \qquad R_P = R_{1|\varnothing} + R_{1|2}, \qquad \mathcal{R} = \tfrac{1}{4}(R_E + R_P). \tag{8}$$

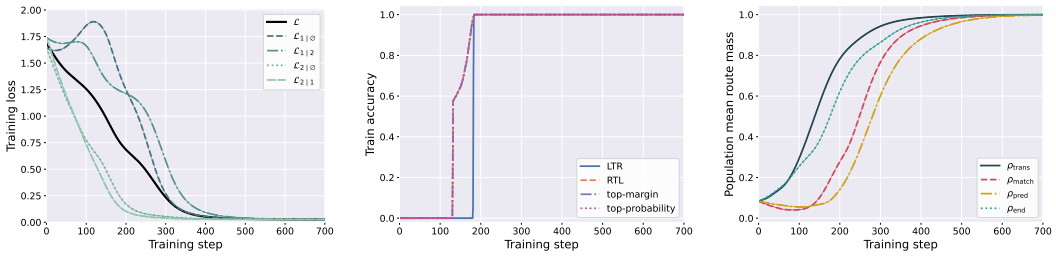
For $X \in \{E, P\}$, let $\tau_X(\eta) = \inf\{t \geq 0 : R_X(t) \leq \eta\}$, with $\inf \varnothing = \infty$.

Theorem 3.1 (Local endpoint-first loss separation)**.** *The paired losses are equal at initialization, and*

$$\dot{\mathcal{L}}_{1|\varnothing}(0) - \dot{\mathcal{L}}_{2|\varnothing}(0) > 0, \qquad \dot{\mathcal{L}}_{1|2}(0) - \dot{\mathcal{L}}_{2|1}(0) > 0. \tag{9}$$

Hence $\mathcal{L}_{2|\varnothing}(t) < \mathcal{L}_{1|\varnothing}(t)$ and $\mathcal{L}_{2|1}(t) < \mathcal{L}_{1|2}(t)$ for all sufficiently small positive t. Moreover, there is $h_E > 0$ such that, for $0 < h < h_E$ and $\eta = R_E(0) - h$, $t_E := \tau_E(\eta) < \tau_P(\eta)$, $t_E < \infty$.

The endpoint's initial lead motivates preparing it before analyzing predecessor learning; Figure 3a provides the corresponding joint-training loss trajectory.



(a) Mask-conditioned losses (b) Decoding accuracy (c) Reverse-circuit route mass

Figure 3: Empirical joint-training chronology on the terminal-pair task. (a) Evolution of the four mask-conditioned losses. (b) Decoding accuracy under fixed and confidence-based orders. (c) Formation of the reverse-reasoning routes, where $\rho_{\text{trans}} = Q_{a_2,q_t}$, $\rho_{\text{match}} = S^{(2)}_{a_1,d^\star}$, $\rho_{\text{pred}} = S^{(2)}_{a_1,d^\star} S^{(1)}_{d^\star,s^\star}$, and $\rho_{\text{end}} = \sum_{k\in\mathcal{A}} S^{(2)}_{a_2,k} S^{(1)}_{k,q_t}$.

3.2 Stage-Wise Predecessor Learning

This ordering motivates, but does not supply, the transition state. We accordingly prescribe entry readiness and endpoint preservation while maintaining symmetry-fixed, row-centered weights and Layer-2 pair blindness throughout the staged phase.

Stage I: endpoint preparation. Preparation supplies a finite endpoint-ready state at $t_0 \geq t_E$. At this transition, choose finite constants $a, b > 0$ and fix

$$\mathbf{W}_c^{(1)} = -a \begin{pmatrix} \mathbf{I}_K & \mathbf{1}_K \\ \mathbf{0}^\top & 1 \end{pmatrix} \mathbf{C}_D \qquad \mathbf{W}_c^{(2)} = b \begin{pmatrix} \mathbf{I}_K & \mathbf{0} \\ \mathbf{0}^\top & 0 \end{pmatrix} \mathbf{C}_D \tag{10}$$

where $\mathbf{C}_D = \mathbf{I}_D - D^{-1}\mathbf{1}_D\mathbf{1}_D^\top$. Layer 1 suppresses same-content and mask competitors relative to different-node keys; Layer 2 favors equality matching.[2] All entry conditions are evaluated after this choice. Write $v^\star = [\mathbf{x}_0]_{a_2}$ and use

$$C_1 := \frac{\alpha(K-1)}{e^\alpha + K - 1}, \qquad \delta_L := \frac{1}{4(L-2)}, \qquad C_2 := \frac{4}{C_1\delta_L}. \tag{11}$$

Corollary 3.2 (Endpoint output-value readiness)**.** *Fix either endpoint mask state* $2 \mid \varnothing$ *or* $2 \mid 1$ *and evaluate the forward pass in that state. If* $R_E(t) \leq \eta$*, then* $H^{(2)}_{a_2,v^\star}(t) \geq 1 - \eta/C_1$.

To identify the underlying copying route on the both-masked input, Lemma B.2 gives $1 - Q_{a_2,q_t} \leq 3\varepsilon_E + e^{B_{\text{self}} - a}$ whenever the endpoint output impurity is at most ε_E, with B_{self} as defined there. Choosing $a \geq B_{\text{self}} + \log(1/\varepsilon_E)$ thus gives $1 - Q_{a_2,q_t} \leq 4\varepsilon_E$ without using predecessor risk.

Lemma 3.3 (Finite-time carrier extraction)**.** *For the both-masked input, the carrier set* $\mathcal{A}$ *satisfies*

$$\sum_{k\in\mathcal{A}} S^{(2)}_{a_2,k} \geq \sum_{k\in\mathcal{A}} S^{(2)}_{a_2,k} S^{(1)}_{k,q_t} \geq 1 - \frac{1 - Q_{a_2,q_t}}{\varepsilon}. \tag{12}$$

In particular, $\mathcal{A}$ *is nonempty whenever* $1 - Q_{a_2,q_t} < \varepsilon$.

The carrier bound hence identifies the endpoint computation that Stage II is prescribed to retain.

Stage II: predecessor learning. From t_0 onward, hold the content blocks fixed and freeze the stipulated endpoint-related degrees of freedom. Let $\mathcal{U}_\ell$ be the remaining trainable positional subspaces. They retain every unfrozen symmetry-allowed coordinate. With $\mathcal{J}_P = \frac{1}{2}(\mathcal{L}_{1|\varnothing} + \mathcal{L}_{1|2})$, the flow is $\dot{W}_p^{(\ell)} = -\operatorname{Proj}_{\mathcal{U}_\ell} \nabla_{W_p^{(\ell)}} \mathcal{J}_P$, $\ell = 1, 2$.

With endpoint copying retained, we next ask what activates the predecessor route. Fix $e \in \{1 \mid \varnothing, 1 \mid 2\}$; forward quantities in the following definitions are evaluated in that predecessor state. The

[2] These content choices lie within the symmetry-fixed, row-centered family and leave positional retrieval routes unspecified.

endpoint mass already present in the predecessor query is $b_e := H^{(1)}_{a_1,v^\star}$. This measures readiness for matching the terminal destination. The route also needs exposure to graph destinations in Layer 2 and graph sources in Layer 1:

$$\chi := \sum_r S^{(2)}_{a_1,d_r}, \qquad \zeta := \sum_r S^{(1)}_{d^\star,s_r}. \tag{13}$$

Let $\Delta^{(2)}_c := W^{(2)}_{c,ii} - W^{(2)}_{c,ij}$ for distinct node identities i, j measure content matching, and let $\gamma_{ds} := W^{(1)}_{p,d_r,s_r} - W^{(1)}_{p,d_r,s_{r'}}$ for $r \neq r'$ measure preference for the paired source. Symmetry makes these contrasts independent of the chosen identities and edges; the selected content gives $\Delta^{(2)}_c = \alpha_2 - \beta_2 = b > 0$.

Define $\kappa = S^{(2)}_{a_2,a_1}$ using the both-masked input only. Lemma B.3 gives $b_e \geq 1 - \varepsilon_E/\kappa$ in both predecessor states. For $b_\star \in (1/2, 1)$, $\varepsilon_E \leq \kappa(1 - b_\star)$ yields $b_e \geq b_\star > 1/2$; finite softmax makes $\kappa > 0$ but does not itself supply this relative-accuracy condition.

The reverse route is controlled by two factors: destination matching in Layer 2 and paired-source selection in Layer 1. Define

$$g^{\text{match}}_e := \left[1 + \sum_{r \neq r_\star} \exp\{-\Delta^{(2)}_c (b_e - H^{(1)}_{a_1,[\mathbf{x}_0]_{d_r}})\}\right]^{-1}, \quad g^{\text{src}} := \frac{\zeta}{1 + (m-1)e^{-\gamma_{ds}}}, \tag{14}$$

and let

$$\underline{r}_e := \frac{\chi\zeta}{\left[[1 + (m-1)e^{-\Delta^{(2)}_c(2b_e-1)}][1 + (m-1)e^{-\gamma_{ds}}]\right]} \tag{15}$$

Proposition 3.4 (Reverse-route certificate)**.** *In either predecessor mask state $e \in \{1 \mid \varnothing, 1 \mid 2\}$, $S^{(2)}_{a_1,d^\star}/\chi = g^{\text{match}}_e$, and $S^{(1)}_{d^\star,s^\star} = g^{\text{src}}$. Hence the terminal reverse-route mass $r_e := S^{(2)}_{a_1,d^\star} S^{(1)}_{d^\star,s^\star}$ satisfies $r_e \geq \underline{r}_e$. Whenever $\underline{r}_e > 1/2$, the predecessor prediction is uniquely correct.*

In consequence, destination matching and paired-source selection are distinct from pool exposure $\chi\zeta$. Query readiness and a sufficiently large fixed b control matching, enabling the descent test below.

From query readiness to a finite risk decrease. Use the entry conditions of Lemma B.4: $b_e(t_0) \geq b_\star > b_0 > 1/2$ for both predecessor states, its paired-source descent direction remains trainable, and b satisfies its finite-scale bound. The lemma gives $c_P, \tau > 0$ with $\dot{R}_P \leq -c_P$ on $[t_0, t_0 + \tau]$. Every threshold satisfying

$$0 < R_P(t_0) - \eta < \min\{R_P(t_0)/2, c_P\tau\} \tag{16}$$

is reached at $t_P := \inf\{t \geq t_0 : R_P(t) \leq \eta\} \leq t_0 + [R_P(t_0) - \eta]/c_P$. For the following certificate, fix such a staged threshold and prescribe $R_E(t) \leq \eta$ on $[t_0, t_P]$. This threshold need not coincide with the early joint-training threshold.

3.3 Reverse-Circuit Certification at the Attained Risk

The reduced predecessor risk and retained endpoint bound now jointly certify both routes. Set $\rho = \eta/2$ and define $t_\rho := \inf\{t \geq t_0 : \mathcal{R}(t) \leq \rho\}$, and $\inf \varnothing := \infty$.

Theorem 3.5 (Finite-time reverse-circuit formation)**.** *We have $t_\rho \leq t_P < \infty$. At $t = t_\rho$ and $t = t_P$, every both-masked population input satisfies $S^{(2)}_{a_1,d^\star}(t) S^{(1)}_{d^\star,s^\star}(t) \geq 1 - C_2\rho$, and $Q_{a_2,q_t}(t) \geq 1 - C_2\rho$. For $0 < \varepsilon < 1$, the attained carrier set $\mathcal{A} = \mathcal{A}_\varepsilon(t; \widetilde{\mathbf{x}})$ also satisfies*

$$\sum_{k \in \mathcal{A}} S^{(2)}_{a_2,k}(t) \geq \sum_{k \in \mathcal{A}} S^{(2)}_{a_2,k}(t) S^{(1)}_{k,q_t}(t) \geq 1 - \frac{C_2\rho}{\varepsilon}, \qquad S^{(1)}_{k,q_t}(t) \geq 1 - \varepsilon \qquad \forall k \in \mathcal{A}. \tag{17}$$

The risk identity gives $\mathcal{R}(t_P) = (R_E(t_P) + R_P(t_P))/4 \leq \eta/2$, so the theorem certifies both predecessor-retrieval steps and endpoint transport through carriers. Each predecessor factor has the same lower bound, and $\mathcal{A} \neq \varnothing$ when $C_2\rho < \varepsilon$. In particular,

$$0 < \rho \leq \frac{C_1\delta_L}{4}\varepsilon^2 \tag{18}$$

guarantees predecessor-route mass at least $1-\varepsilon^2$ and endpoint-route mass through $\mathcal{A}$ at least $1-\varepsilon$. This route tolerance additionally requires the attained threshold to satisfy the displayed bound; Figure 3c tracks these same route masses under joint training.

4 From Learned Computation to Decoding

Section 3 certifies internal reverse retrieval, but not an output order. To interpret the joint-training patterns in Figure 3, we therefore separate endpoint-reveal benefits, confidence-based selection, and order independence at vanishing risk.

Decoding rules and sequence accuracy. LTR resolves (a_1, a_2); RTL resolves (a_2, a_1). Both greedily commit a highest-probability token at each step without remasking. For a masked answer a_i, let $p_i(v)$ be its current token probability, with largest two values $p_i^{(1)} \geq p_i^{(2)}$. Top-probability and top-margin select the position maximizing

$$c_i^{\mathrm{prob}} := \max_{v\in\mathcal{V}} p_i(v), \qquad c_i^{\mathrm{margin}} := p_i^{(1)} - p_i^{(2)}. \tag{19}$$

Fix a tie-breaking rule, and let $\mathsf{E}(e;t)$ denote the event of greedy correctness in problem $e \in \mathcal{E}_4$ with ground-truth reveals. A correct first token produces exactly that revealed state, giving the success identities

$$\mathrm{Acc}_{\mathrm{LTR}}(t) = \mathbb{P}\big(\mathsf{E}(1 \mid \varnothing; t) \cap \mathsf{E}(2 \mid 1; t)\big), \qquad \mathrm{Acc}_{\mathrm{RTL}}(t) = \mathbb{P}\big(\mathsf{E}(2 \mid \varnothing; t) \cap \mathsf{E}(1 \mid 2; t)\big).$$

4.1 From Endpoint Preparation to Reveal Bootstrap

Before endpoint readiness. The early endpoint-first loss separation can precede any decoding success (Corollary C.1). In light of this, we ask how revealing the prepared endpoint changes the Stage II predecessor query. Since the endpoint already appears at q_t, reveal changes access rather than task information.

Compare both answers masked (unprimed) with $a_2 = v^\star$ revealed and a_1 masked (primed), at the same parameters. With $y_1 = [\mathbf{x}_0]_{a_1}$, define $M := H^{(2)}_{a_1,y_1} - \max_{v\in\mathcal{V}\setminus\{y_1\}} H^{(2)}_{a_1,v}$ and $r := S^{(2)}_{a_1,d^\star} S^{(1)}_{d^\star,s^\star}$. The primed formulas define M', r'. Write $\chi = \sum_{j=1}^m S^{(2)}_{a_1,d_j}$, and similarly χ', for destination-pool exposure.

Lemma 4.1 (Reveal improves within-pool destination matching)**.** *At any finite state in the Stage II content and symmetry setting of Section 3.2, endpoint reveal satisfies* $H^{(1)\prime}_{a_1,v^\star} > H^{(1)}_{a_1,v^\star}$, *and* $S^{(2)\prime}_{a_1,d^\star}/\chi' > S^{(2)}_{a_1,d^\star}/\chi$, *while* $S^{(1)\prime}_{d^\star,:} = S^{(1)}_{d^\star,:}$.

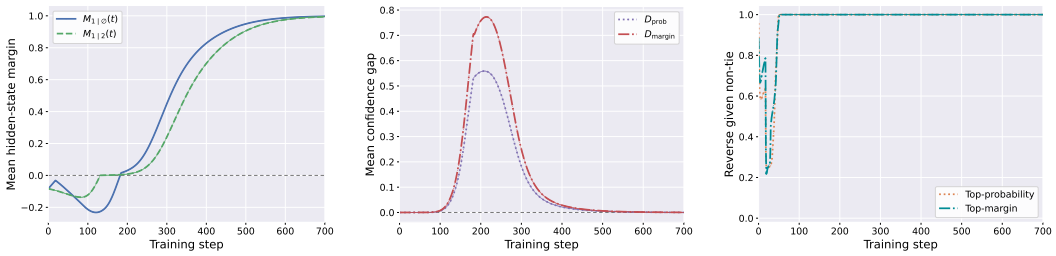
Yet changes in pool exposure or other paths can offset better matching. Let $\varepsilon_{\mathrm{int}} \geq 0$ be the maximum adverse change, over wrong tokens, in correct-versus-wrong output mass contributed by all paths except $a_1 \to d^\star \to s^\star$. Lemma C.3 gives $M' - M \geq r' - r - \varepsilon_{\mathrm{int}}$, translating net route gain into a margin bound.

Theorem 4.2 (Conditional endpoint-reveal bootstrap)**.** *Consider a continuous, symmetry-preserving Stage II trajectory with* $M(t_0) < 0$ *and finite first contact* $t_\star := \inf\{t \geq t_0 : M(t) \geq 0\} < \infty$. *Assume that, at* $t_\star$, *the both-masked endpoint value satisfies* $H^{(2)}_{a_2,v^\star} > 1/2$ *and the reveal gain satisfies* $r' - r > \varepsilon_{\mathrm{int}}$. *Then there is* $0 < \delta < t_\star - t_0$ *such that* $M'(t) > 0 > M(t)$ *and* $\mathrm{Acc}_{\mathrm{RTL}}(t) = 1 > 0 = \mathrm{Acc}_{\mathrm{LTR}}(t)$ *for all* $t \in I_{\mathrm{RTL}} = (t_\star - \delta, t_\star)$.

With the endpoint correct, $M' > 0 > M$ makes RTL succeed while LTR still fails. This connects the margin comparison in Figure 4a to the joint-training accuracy comparison in Figure 3b.

4.2 When Confidence Selects Reverse Decoding

To exploit this advantage, confidence-based decoding must select the endpoint first. Section 3's reverse-route obstruction lets us compare confidences even when the predecessor is incorrect. On the both-masked input, write $\delta_Q = 1 - Q_{a_2,q_t}$ and retain r above and δ_L from equation 11.



(a) Endpoint-reveal margins (b) Confidence gaps (c) Reverse choice among non-ties

Figure 4: Empirical decoding dynamics under joint training. (a) Predecessor margins with and without endpoint reveal (Theorem 4.2). (b) Confidence differences between a_2 and a_1 for top-probability and top-margin decoding (Proposition 4.3 and Theorem 4.4). (c) Fraction of non-tied decisions that select the reverse order; this measures order preference, not sequence accuracy.

Proposition 4.3 (Circuit separation implies reverse confidence)**.** *At a symmetry-fixed, Layer-2 pair-blind state, suppose the both-masked forward pass satisfies* $2\delta_Q < \min\{r, \delta_L(1-r)\}$*. Then* $c_2^{\text{prob}} > c_1^{\text{prob}}$ *and* $c_2^{\text{margin}} > c_1^{\text{margin}}$ *on every population input: both rules strictly choose* a_2 *first.*

The criterion compares endpoint transport error with predecessor route mass and its residual deficit. Since a positive confidence gap selects RTL on each input, Figures 4b and 4c separately track gaps and non-tied reverse choices. When the criterion also holds at a time in I_{RTL}, both adaptive rules exploit the successful RTL rollout.

4.3 Final-Stage Order Independence and Confidence Agreement

To ask whether this advantage persists, consider continuation beyond local Stage II attainment, allowing the content freeze to be released while preserving symmetry. On the both-masked input, write $D_{\text{prob}} = c_2^{\text{prob}} - c_1^{\text{prob}}$ and $D_{\text{margin}} = c_2^{\text{margin}} - c_1^{\text{margin}}$.

Theorem 4.4 (Order independence and confidence agreement)**.** *Suppose a symmetry-preserving continuation satisfies: for every sufficiently small* $\eta > 0$*, there is a finite* $t_P(\eta)$ *with* $R_E(t), R_P(t) \le \eta$ *for all* $t \ge t_P(\eta)$*. For each such* $\eta < C_1/2$*, every monotone greedy order* o *has* $\text{Acc}_o(t) = 1$ *for all* $t \ge t_P(\eta)$*. Moreover, the both-masked confidence gaps satisfy* $|D_{\text{prob}}(t)| \le (2\alpha/C_1)\mathcal{R}(t) \to 0$ *and* $|D_{\text{margin}}(t)| \le (4\alpha/C_1)\mathcal{R}(t) \to 0$ *as* $t \to \infty$*.*

Under this low-risk continuation, accuracy becomes order-independent even if confidence still favors RTL. The bounds thus concern gap magnitudes in Figure 4b, not the reverse-choice fraction in Figure 4c: vanishing gaps need not yield an exact finite-time tie.

5 Conclusion

Masked diffusion models permit flexible generation orders, yet an order helps only if training produces a computation that exploits the context it reveals. We studied how this arises on a path-star planning task with a two-layer Transformer, using a local endpoint-first asymmetry to motivate a stage-wise analysis that separates endpoint preparation from predecessor learning. Within this framework, the attained risk certifies a reverse-reasoning circuit that matches the destination and retrieves its predecessor. This computation determines which order is effective: revealing the prepared endpoint supplies an explicit destination cue, opening a training window in which right-to-left decoding succeeds while left-to-right decoding still fails, and confidence-based decoding selects this reverse order under a circuit-separation condition. As all mask-conditioned risks vanish, every greedy order becomes correct, so the accuracy advantage disappears even if confidence still favors reverse decoding. Decoding-order choice therefore matters most in an intermediate training regime where the endpoint is ready but predecessor retrieval is incomplete, suggesting that decoding should be understood jointly with learning. Extending this conditional analysis to the full joint dynamics and longer reasoning chains remains future work.

AI USE STATEMENT

We used generative AI tools to improve writing and readability, including wording, formatting, and notation checks. AI was also used to draft text based on structures specified by the authors. The organization, intended message, and characterization of the paper's contributions were determined by the authors. For literature-related tasks, generative AI was used to search for related work and summarize existing literature and proofs. The authors independently selected and checked the works that were ultimately referenced. For mathematical claims, generative AI assisted with gradient-flow derivations, inequality calculations, and parts of their proofs. The main mathematical ideas and proof strategies were developed by the authors, and all AI-assisted derivations were verified step by step by the authors. We take responsibility for the final content of this work.

REFERENCES

Jacob Austin, Daniel D. Johnson, Jonathan Ho, Daniel Tarlow, and Rianne van den Berg. Structured denoising diffusion models in discrete state-spaces. In *NeurIPS*, 2021.

Gregor Bachmann and Vaishnavh Nagarajan. The pitfalls of next-token prediction. In *ICML*, 2024.

Heli Ben-Hamu, Itai Gat, Daniel Severo, Niklas S. Nolte, and Brian Karrer. Accelerated sampling from masked diffusion models via entropy bounded unmasking. In *NeurIPS*, 2025.

Lukas Berglund, Meg Tong, Maximilian Kaufmann, Mikita Balesni, Asa Cooper Stickland, Tomasz Korbak, and Owain Evans. The reversal curse: LLMs trained on "A is B" fail to learn "B is A". In *ICLR*, 2024.

Alberto Bietti, Vivien Cabannes, Diane Bouchacourt, Hervé Jégou, and Léon Bottou. Birth of a transformer: A memory viewpoint. In *NeurIPS*, 2023.

Tony Bonnaire, Raphaël Urfin, Giulio Biroli, and Marc Mézard. Why diffusion models don't memorize: The role of implicit dynamical regularization in training. In *NeurIPS*, 2025.

Blake Bordelon, Hamza Chaudhry, and Cengiz Pehlevan. Infinite limits of multi-head Transformer dynamics. In *NeurIPS*, 2024.

Sitan Chen, Kevin Cong, and Jerry Li. Optimal inference schedules for masked diffusion models. In *COLT*, 2026a.

Siyu Chen, Heejune Sheen, Tianhao Wang, and Zhuoran Yang. Unveiling induction heads: Provable training dynamics and feature learning in transformers. In *NeurIPS*, 2024.

Zheng-An Chen, Pengxiao Lin, Zhi-Qin John Xu, and Tao Luo. Focus and dilution: The multi-stage learning process of attention. In *ICML*, 2026b.

Satoshi Hayakawa, Yuhta Takida, Masaaki Imaizumi, Hiromi Wakaki, and Yuki Mitsufuji. Demystifying MaskGIT sampler and beyond: Adaptive order selection in masked diffusion. *Transactions on Machine Learning Research*, 2026. URL https://openreview.net/forum?id=mKlW68i2Ig.

Chunsan Hong, Seonho An, Min-Soo Kim, and Jong Chul Ye. Improving discrete diffusion unmasking policies beyond explicit reference policies. In *ICLR*, 2026.

Jianhao Huang and Baharan Mirzasoleiman. Tuning the implicit regularizer of masked diffusion language models: Enhancing generalization via insights from k-parity. In *ICML*, 2026.

Jianhao Huang, Zixuan Wang, and Jason D. Lee. Transformers learn to implement multi-step gradient descent with chain of thought. In *ICLR*, 2025.

Jianhao Huang, Zhanpeng Zhou, Renqiu Xia, Baharan Mirzasoleiman, Weijie Su, and Wei Huang. How transformers learn to plan via multi-token prediction. In *COLM*, 2026.

Ryotaro Kawata, Yujin Song, Alberto Bietti, Naoki Nishikawa, Taiji Suzuki, Samuel Vaiter, and Denny Wu. From shortcut to induction head: How data diversity shapes algorithm selection in transformers. In *NeurIPS*, 2025.

Jaeyeon Kim, Kulin Shah, Vasilis Kontonis, Sham M. Kakade, and Sitan Chen. Train for the worst, plan for the best: Understanding token ordering in masked diffusions. In *ICML*, 2025.

Jaeyeon Kim, Jonathan Geuter, David Alvarez-Melis, Sham Kakade, and Sitan Chen. Stop training for the worst: Progressive unmasking accelerates masked diffusion training. In *ICML*, 2026.

Juno Kim and Taiji Suzuki. Transformers provably solve parity efficiently with chain of thought. In *ICLR*, 2025.

Pengxiang Li, Yefan Zhou, Dilxat Muhtar, Lu Yin, Shilin Yan, Li Shen, Yi Liang, Soroush Vosoughi, and Shiwei Liu. Diffusion language model knows the answer before it decodes. In *ICLR*, 2026.

Aaron Lou, Chenlin Meng, and Stefano Ermon. Discrete diffusion modeling by estimating the ratios of the data distribution. In *ICML*, 2024.

Hancheng Min, Enrique Mallada, and René Vidal. Early neuron alignment in two-layer ReLU networks with small initialization. In *ICLR*, 2024.

Eshaan Nichani, Alex Damian, and Jason D. Lee. How transformers learn causal structure with gradient descent. In *ICML*, 2024.

Shen Nie, Fengqi Zhu, Chao Du, Tianyu Pang, Qian Liu, Guangtao Zeng, Min Lin, and Chongxuan Li. Scaling up masked diffusion models on text. In *ICLR*, 2025a.

Shen Nie, Fengqi Zhu, Zebin You, Xiaolu Zhang, Jingyang Ou, Jun Hu, Jun Zhou, Yankai Lin, Ji-Rong Wen, and Chongxuan Li. Large language diffusion models. In *NeurIPS*, 2025b.

Catherine Olsson, Nelson Elhage, Neel Nanda, Nicholas Joseph, Nova DasSarma, Tom Henighan, Ben Mann, Amanda Askell, Yuntao Bai, Anna Chen, Tom Conerly, Dawn Drain, Deep Ganguli, Zac Hatfield-Dodds, Danny Hernandez, Scott Johnston, Andy Jones, Jackson Kernion, Liane Lovitt, Kamal Ndousse, Dario Amodei, Tom Brown, Jack Clark, Jared Kaplan, Sam McCandlish, and Chris Olah. In-context learning and induction heads. *Transformer Circuits Thread*, 2022. URL https://transformer-circuits.pub/2022/in-context-learning-and-induction-head.

Fred Zhangzhi Peng, Zachary Bezemek, Sawan Patel, Jarrid Rector-Brooks, Sherwood Yao, Avishek Joey Bose, Alexander Tong, and Pranam Chatterjee. Path planning for masked diffusion model sampling. *arXiv preprint arXiv:2502.03540*, 2025.

Fred Zhangzhi Peng, Zachary Bezemek, Jarrid Rector-Brooks, Shuibai Zhang, Anru R. Zhang, Michael M. Bronstein, Avishek Joey Bose, and Alexander Tong. Planner aware path learning in diffusion language models training. In *ICLR*, 2026.

Scott Pesme and Nicolas Flammarion. Saddle-to-saddle dynamics in diagonal linear networks. In *NeurIPS*, 2023.

Yi Ren and Danica J. Sutherland. Learning dynamics of LLM finetuning. In *ICLR*, 2025.

Subham Sekhar Sahoo, Marianne Arriola, Yair Schiff, Aaron Gokaslan, Edgar Marroquin, Justin T Chiu, Alexander Rush, and Volodymyr Kuleshov. Simple and effective masked diffusion language models. In *NeurIPS*, 2024.

Abulhair Saparov and He He. Language models are greedy reasoners: A systematic formal analysis of chain-of-thought. In *ICLR*, 2023.

Jiaxin Shi, Kehang Han, Zhe Wang, Arnaud Doucet, and Michalis K. Titsias. Simplified and generalized masked diffusion for discrete data. In *NeurIPS*, 2024.

Itamar Trainin, Shauli Ravfogel, Omri Abend, and Amir Feder. Discrete diffusion models exploit asymmetry to solve lookahead planning tasks. *arXiv preprint arXiv:2602.19980*, 2026.

Binxu Wang and Cengiz Pehlevan. An analytical theory of spectral bias in the learning dynamics of diffusion models. In *NeurIPS*, 2025.

Jiawei Xu, Minghui Liu, Aakriti Agrawal, Yifan Chen, and Furong Huang. Scheduling thoughts: Learning the order of thought in diffusion language models. In *ICML*, 2026.

Jiacheng Ye, Jiahui Gao, Shansan Gong, Lin Zheng, Xin Jiang, Zhenguo Li, and Lingpeng Kong. Beyond autoregression: Discrete diffusion for complex reasoning and planning. In *ICLR*, 2025.

Taesun Yeom, Taehyeok Ha, and Jaeho Lee. Over-alignment vs over-fitting: The role of feature learning strength in generalization. In *ICML*, 2026.

Shaorong Zhang, Longxuan Yu, Rob Brekelmans, Luhan Tang, Salman Asif, and Greg Ver Steeg. Generation order and parallel decoding in masked diffusion models: An information-theoretic perspective. *arXiv preprint arXiv:2602.00286*, 2026.

Yedi Zhang, Aaditya K. Singh, Peter E. Latham, and Andrew M. Saxe. Training dynamics of in-context learning in linear attention. In *ICML*, 2025.

Kaiwen Zheng, Yongxin Chen, Hanzi Mao, Ming-Yu Liu, Jun Zhu, and Qinsheng Zhang. Masked diffusion models are secretly time-agnostic masked models and exploit inaccurate categorical sampling. In *ICLR*, 2025.

APPENDIX CONTENTS

A Proofs for Section 2 **14**

A.1 Data Symmetry 14

A.2 Gradient-Flow Invariance 15

A.3 Exact Fixed Weight Structure 16

A.4 Layer-2 Pair Blindness 19

B Proofs for Section 3 **20**

B.1 Exact gradients and projections 20

B.2 Proof of the local endpoint-first separation 22

B.3 Stage I: endpoint provenance, carriers, and query transfer 23

B.4 Gate identities 25

B.5 Stage II: predecessor descent and finite-interval attainment 25

B.6 Proofs for the reverse-circuit certificate 27

C Proofs for Section 4 **30**

C.1 Decoding Strategies 30

C.2 Before Endpoint Readiness 30

C.3 Stage II: Endpoint-Reveal Bootstrap 31

C.4 From Circuit Separation to Reverse Confidence 34

C.5 Final-Stage Order Independence and Confidence Agreement 34

D Experimental Details **36**

D.1 Model Training Setup 36

D.2 $G(R, L)$ Results with Minibatch Setting 36

A PROOFS FOR SECTION 2

A.1 DATA SYMMETRY

Throughout, fix $R \geq 2$ and $L \geq 3$, with $K = 1 + R(L-1)$, $m = R(L-1)$, $D = K + 1$, and $T = 2m + 4$. We first show that vertex relabeling and directed edge serialization preserve the prediction problem. Their joint orbit then accounts for non-target arm permutations and connects the population symmetries to gradient flow.

Proposition A.1 (Vertex-relabeling symmetry)**.** *For every $\tau \in S_K$, simultaneously replacing each node token $v \in \mathcal{V}$ by $\tau(v)$ maps the population $\mathcal{D}$ bijectively onto itself. If $\mathbf{x}_0 \sim \mathrm{Unif}(\mathcal{D})$, then $\tau \cdot \mathbf{x}_0 \sim \mathrm{Unif}(\mathcal{D})$. The population distinguishes equality relations between node tokens, but not the node identities themselves.*

Proof. For $\tau \in S_K$, let $\tau(\mathcal{E}) = \{(\tau(v), \tau(w)) : (v, w) \in \mathcal{E}\}$. Since τ acts on node identities whereas π acts only on the serialization order of directed edge pairs, the two actions commute. For every $\mathbf{x}_0 \in \mathcal{D}$,

$$\tau \cdot \mathbf{x}_0 = \tau \cdot [\pi(\mathcal{E}) \mid u, v_{L-1} \mid v_{L-2}, v_{L-1}] \tag{20}$$

$$= [\pi(\tau(\mathcal{E})) \mid \tau(u), \tau(v_{L-1}) \mid \tau(v_{L-2}), \tau(v_{L-1})] \in \mathcal{D}. \tag{21}$$

Applying the same argument to τ^{-1} shows that $\mathbf{x}_0 \mapsto \tau \cdot \mathbf{x}_0$ is a bijection of $\mathcal{D}$. Reindexing the finite uniform average by $\mathbf{x}_0' = \tau \cdot \mathbf{x}_0$ gives, for any $f : \mathcal{D} \to \mathbb{R}$,

$$\mathbb{E}_{\mathbf{x}_0 \sim \mathrm{Unif}(\mathcal{D})}[f(\tau \cdot \mathbf{x}_0)] = \frac{1}{|\mathcal{D}|} \sum_{\mathbf{x}_0 \in \mathcal{D}} f(\tau \cdot \mathbf{x}_0) \tag{22}$$

$$= \frac{1}{|\mathcal{D}|} \sum_{\mathbf{x}_0' \in \mathcal{D}} f(\mathbf{x}_0') \tag{23}$$

$$= \mathbb{E}_{\mathbf{x}_0 \sim \mathrm{Unif}(\mathcal{D})}[f(\mathbf{x}_0)], \tag{24}$$

which proves invariance of the uniform population distribution. □

Remark. Vertex relabeling removes dependence on arbitrary node names, leaving equality relations as the content information retained by the symmetry reduction.

Proposition A.2 (Directed edge-serialization symmetry)**.** *For every $\sigma \in S_m$, simultaneously permuting the directed edge slots according to $s_r \mapsto s_{\sigma(r)}$ and $d_r \mapsto d_{\sigma(r)}$, while fixing q_s, q_t, a_1, a_2, defines a bijection of $\mathcal{D}$. In particular, if $\mathbf{x}_0 \sim \mathrm{Unif}(\mathcal{D})$, then $\sigma \cdot \mathbf{x}_0 \sim \mathrm{Unif}(\mathcal{D})$. The action preserves the source–destination orientation within every directed edge and therefore does not exchange s_r and d_r. Under the induced diagonal action $(r, r') \mapsto (\sigma(r), \sigma(r'))$ on ordered edge-index pairs, there are exactly two orbits, $\mathcal{O}_= = \{(r, r) : r \in [m]\}$ and $\mathcal{O}_\neq = \{(r, r') : r \neq r'\}$.*

Proof. Since π ranges over all elements of S_m, left composition by σ permutes the admissible serializations, with inverse given by left composition with σ^{-1}. Hence, for every $\mathbf{x}_0 \in \mathcal{D}$,

$$\sigma \cdot \mathbf{x}_0 = \sigma \cdot [\pi(\mathcal{E}) \mid u, v_{L-1} \mid v_{L-2}, v_{L-1}] \tag{25}$$

$$= [(\sigma \circ \pi)(\mathcal{E}) \mid u, v_{L-1} \mid v_{L-2}, v_{L-1}] \in \mathcal{D}. \tag{26}$$

Thus $\mathbf{x}_0 \mapsto \sigma \cdot \mathbf{x}_0$ is a bijection of $\mathcal{D}$ and preserves its uniform distribution. Because complete directed edge pairs are permuted, their source and destination roles are preserved. For the induced action on ordered edge-index pairs, bijectivity preserves whether the two indices are equal. For arbitrary $r \neq r'$ and $k \neq k'$, there exists $\sigma \in S_m$ with $\sigma(r) = k$ and $\sigma(r') = k'$. Likewise, for arbitrary $r, k \in [m]$, one may choose $\sigma(r) = k$. Thus all diagonal pairs form one orbit and all off-diagonal pairs form the other, yielding exactly $\mathcal{O}_=$ and $\mathcal{O}_\neq$. □

Remark. Edge serialization preserves source–destination orientation. The positional reduction therefore retains graph roles and same-edge incidence, rather than an order on the edge list.

Since the node-relabeling and edge-serialization actions commute, we can combine them to describe the entire task population as one orbit.

Proposition A.3 (Orbit representation of the task population)**.** *Let $\mathcal{G} = S_K \times S_m$ act on clean sequences through the vertex-relabeling and directed edge-serialization actions of Propositions A.1 and A.2. For any reference sequence $\mathbf{x}^\circ \in \mathcal{D}$, the population is the single orbit $\mathcal{D} = \mathcal{G} \cdot \mathbf{x}^\circ$, and $\mathrm{Stab}_{\mathcal{G}}(\mathbf{x}^\circ) \cong S_{R-1}$, corresponding to permutations of the non-target arms. Hence, by the orbit–stabilizer theorem, $|\mathcal{D}| = K!m!/(R-1)!$. Every sequence has exactly $(R-1)!$ preimages under the orbit map, so for every $f : \mathcal{D} \to \mathbb{R}$, $\mathbb{E}_{\mathbf{x}_0 \sim \mathrm{Unif}(\mathcal{D})}[f(\mathbf{x}_0)] = \frac{1}{|S_K||S_m|} \sum_{\tau \in S_K} \sum_{\sigma \in S_m} f((\tau, \sigma) \cdot \mathbf{x}^\circ)$.*

Proof of Proposition A.3. The two actions commute because vertex relabeling acts on token identities while edge serialization acts on edge-slot positions. Every clean sequence has the same directed R-arm structure and terminal-pair format as $\mathbf{x}^\circ$. A graph isomorphism matching the target arms gives an appropriate vertex relabeling, and an edge permutation then matches the serialization. Hence $\mathcal{D} = \mathcal{G} \cdot \mathbf{x}^\circ$.

To identify the stabilizer, suppose $(\tau, \sigma) \cdot \mathbf{x}^\circ = \mathbf{x}^\circ$. Equality of the edge lists implies that τ is an automorphism of the underlying directed graph. Since σ fixes the four special positions, the suffix forces τ to fix the root and the final two vertices of the target arm. Every graph automorphism permutes complete arms while preserving depth along each arm; fixing the target endpoint therefore fixes the entire target arm. The remaining $R-1$ arms may be permuted arbitrarily. Each such vertex permutation has a unique accompanying edge-slot permutation restoring the original serialization, because all directed edges are distinct. Projection onto the vertex permutation consequently identifies the stabilizer with S_{R-1}.

Thus $|\mathrm{Stab}_{\mathcal{G}}(\mathbf{x}^\circ)| = (R-1)!$, and the orbit–stabilizer theorem gives

$$|\mathcal{D}| = |\mathcal{G} \cdot \mathbf{x}^\circ| = \frac{|\mathcal{G}|}{|\mathrm{Stab}_{\mathcal{G}}(\mathbf{x}^\circ)|} = \frac{|S_K||S_m|}{(R-1)!} = \frac{K!m!}{(R-1)!}. \tag{27}$$

Every fiber of the orbit map $(\tau, \sigma) \mapsto (\tau, \sigma) \cdot \mathbf{x}^\circ$ is a coset of the stabilizer and has size $(R-1)!$. Therefore the uniform group average counts every clean sequence equally and gives the stated uniform population average. For $R = 2$, the stabilizer is trivial, recovering the two-arm formula. □

Remark. The stabilizer accounts for repeated descriptions of the same sequence. Consequently, uniform averaging over relabelings and serializations gives the uniform task population for every $R \geq 2$.

A.2 Gradient-Flow Invariance

The orbit representation now transfers data symmetry to the objective. Gradient equivariance then keeps the zero-initialized flow in the fixed spaces, without determining its learned attention preferences.

Symmetry reduction for the population flow. Consider $\dot{\theta}(t) = -\nabla \mathcal{L}(\theta(t))$ with $\theta(0) = 0$. Let $P_\tau \in \mathbb{R}^{K \times K}$ be the permutation matrix associated with $\tau \in S_K$ and let $C_\tau = \mathrm{diag}(P_\tau, 1) \in \mathbb{R}^{D \times D}$. Let $Q_\sigma \in \mathbb{R}^{T \times T}$ act by $(s_r, d_r) \mapsto (s_{\sigma(r)}, d_{\sigma(r)})$ and fix q_s, q_t, a_1, a_2.

Lemma A.4 (Loss invariance, gradient equivariance, and invariant gradient flow)**.** *For $g = (\tau, \sigma) \in S_K \times S_m$, let ρ_g act on the attention parameters by $\mathbf{W}_c^{(\ell)} \mapsto C_\tau \mathbf{W}_c^{(\ell)} C_\tau^\top$ and $\mathbf{W}_p^{(\ell)} \mapsto Q_\sigma \mathbf{W}_p^{(\ell)} Q_\sigma^\top$. Then $\mathcal{L}(\rho_g(\theta)) = \mathcal{L}(\theta)$ and $\nabla \mathcal{L}(\rho_g(\theta)) = \rho_g(\nabla \mathcal{L}(\theta))$. Consequently, the gradient flow satisfies $\rho_g(\theta(t)) = \theta(t)$ for every $g \in S_K \times S_m$ and $t \geq 0$, so $\mathbf{W}_c^{(\ell)}(t) \in \mathrm{Fix}(S_K)$ and $\mathbf{W}_p^{(\ell)}(t) \in \mathrm{Fix}(S_m)$.*

Proof of Lemma A.4. For arbitrary parameters θ and $g = (\tau, \sigma)$, let $\mathbf{S}_g^{(\ell)}$ and $\mathbf{\Pi}_g$ denote the forward quantities evaluated at $\rho_g(\theta)$ on the transformed masked input. Since node relabeling fixes the mask token and edge serialization fixes the answer positions, masking commutes with the group action, and the corrupted content matrix transforms as $\mathbf{Z}^g = Q_\sigma \mathbf{Z} C_\tau^\top$. The fixed output head satisfies $\mathbf{U} C_\tau = P_\tau \mathbf{U}$, and row-wise softmax is equivariant under row and column permutations. Applying

these identities successively gives

$$\mathbf{S}_g^{(1)} = \mathrm{softmax}\Big(Q_\sigma \left(\mathbf{Z}\mathbf{W}_c^{(1)}\mathbf{Z}^\top + \mathbf{W}_p^{(1)}\right) Q_\sigma^\top\Big) = Q_\sigma \mathbf{S}^{(1)} Q_\sigma^\top, \tag{28}$$

$$\mathbf{S}_g^{(2)} = \mathrm{softmax}\Big(Q_\sigma \mathbf{S}^{(1)} \left(\mathbf{Z}\mathbf{W}_c^{(2)}\mathbf{Z}^\top + \mathbf{W}_p^{(2)}\right) Q_\sigma^\top\Big) = Q_\sigma \mathbf{S}^{(2)} Q_\sigma^\top, \tag{29}$$

$$\mathbf{\Pi}_g = \mathrm{softmax}\Big(Q_\sigma \mathbf{S}^{(2)}\mathbf{S}^{(1)}\mathbf{Z}C_\tau^\top \mathbf{U}^\top\Big) = \mathrm{softmax}\Big(Q_\sigma \mathbf{S}^{(2)}\mathbf{S}^{(1)}\mathbf{Z}\mathbf{U}^\top P_\tau^\top\Big) = Q_\sigma \mathbf{\Pi} P_\tau^\top. \tag{30}$$

The target token at each masked answer position is relabeled by τ, while Q_σ fixes the answer positions. Hence the masked cross-entropy is unchanged:

$$\ell_{\mathrm{ex}}(\rho_g(\theta); g\mathbf{x}_0, M) = -\frac{1}{|M|}\sum_{i\in M} \log\left[Q_\sigma \mathbf{\Pi} P_\tau^\top\right]_{i,\tau(x_{0,i})} \tag{31}$$

$$= -\frac{1}{|M|}\sum_{i\in M} \log \mathbf{\Pi}_{i,x_{0,i}} \tag{32}$$

$$= \ell_{\mathrm{ex}}(\theta; \mathbf{x}_0, M). \tag{33}$$

By Proposition A.3, uniform averaging over $\mathcal{G}$ gives the population expectation despite the nontrivial stabilizer. In particular, the group-average normalization remains $|\mathcal{G}|^{-1} = 1/(K!m!)$. The mask distribution is also unchanged, so

$$\mathcal{L}(\rho_g(\theta)) = \frac{1}{K!m!}\sum_{h\in\mathcal{G}} \mathbb{E}[\ell_{\mathrm{ex}}(\rho_g(\theta); h\cdot\mathbf{x}^\circ, M)] \tag{34}$$

$$= \frac{1}{K!m!}\sum_{h\in\mathcal{G}} \mathbb{E}\big[\ell_{\mathrm{ex}}(\theta; g^{-1}h\cdot\mathbf{x}^\circ, M)\big] \tag{35}$$

$$= \frac{1}{K!m!}\sum_{h'\in\mathcal{G}} \mathbb{E}[\ell_{\mathrm{ex}}(\theta; h'\cdot\mathbf{x}^\circ, M)] \tag{36}$$

$$= \mathcal{L}(\theta). \tag{37}$$

The action ρ_g is orthogonal under the product Frobenius inner product. Differentiating loss invariance in an arbitrary parameter direction h gives

$$\langle \nabla\mathcal{L}(\rho_g(\theta)), h\rangle = D\mathcal{L}(\rho_g(\theta))[h] \tag{38}$$

$$= D\mathcal{L}(\theta)[\rho_g^{-1}(h)] \tag{39}$$

$$= \big\langle \nabla\mathcal{L}(\theta), \rho_g^{-1}(h)\big\rangle \tag{40}$$

$$= \langle \rho_g(\nabla\mathcal{L}(\theta)), h\rangle\,, \tag{41}$$

and hence $\nabla\mathcal{L}(\rho_g(\theta)) = \rho_g(\nabla\mathcal{L}(\theta))$. Finally, define $\theta_g(t) = \rho_g(\theta(t))$. Gradient equivariance and zero initialization give

$$\dot\theta_g(t) = \rho_g(\dot\theta(t)) = -\rho_g(\nabla\mathcal{L}(\theta(t))) = -\nabla\mathcal{L}(\theta_g(t)), \tag{42}$$

$$\theta_g(0) = \rho_g(\theta(0)) = \theta(0). \tag{43}$$

Thus θ_g and θ solve the same initial-value problem, and uniqueness gives $\theta_g(t) = \theta(t)$. Smoothness and the energy bound $\int_0^t \|\dot\theta(s)\|_F^2\, ds \le \mathcal{L}(\theta(0))$ exclude finite-time escape, so the identity holds for all $t \ge 0$. Hence the content and positional trajectories remain in $\mathrm{Fix}(S_K)$ and $\mathrm{Fix}(S_m)$, respectively. □

Remark. The symmetry-fixed family thus describes the actual population trajectory, rather than an additional architectural restriction. Its fixed spaces can now be characterized entry by entry.

A.3 Exact Fixed Weight Structure

Proof of Lemma 2.1. Fix a layer $\ell \in \{1, 2\}$, retaining layer indices on its coefficients to distinguish them from the fixed head scale α. Write $\mathbf{W}_c^{(\ell)}$ in node–mask block form as $\mathbf{W}_c^{(\ell)} = \begin{pmatrix} \mathbf{A} & \mathbf{r} \\ \mathbf{s}^\top & \mu_\ell \end{pmatrix}$.

Since Lemma A.4 gives $C_\tau \mathbf{W}_c^{(\ell)} C_\tau^\top = \mathbf{W}_c^{(\ell)}$ for every $\tau \in S_K$, the node–node block is invariant under simultaneous permutation of its row and column indices, while the node–mask and mask–node blocks are invariant under permutation of their node coordinates:

$$P_\tau \mathbf{A} P_\tau^\top = \mathbf{A}, \qquad P_\tau \mathbf{r} = \mathbf{r}, \qquad \mathbf{s}^\top P_\tau^\top = \mathbf{s}^\top. \tag{44}$$

The action of S_K is transitive on diagonal ordered node pairs, off-diagonal ordered node pairs, and individual node coordinates. Thus, for arbitrary $i, j, k, l \in [K]$ with $i \neq j$ and $k \neq l$,

$$A_{ii} = A_{kk} = \alpha_\ell, \qquad A_{ij} = A_{kl} = \beta_\ell, \qquad r_i = r_k = \rho_\ell, \qquad s_i = s_k = \varsigma_\ell, \tag{45}$$

which gives

$$\mathbf{A} = \alpha_\ell \mathbf{I}_K + \beta_\ell (\mathbf{1}_K \mathbf{1}_K^\top - \mathbf{I}_K), \qquad \mathbf{r} = \rho_\ell \mathbf{1}_K, \qquad \mathbf{s} = \varsigma_\ell \mathbf{1}_K, \tag{46}$$

and hence

$$\mathbf{W}_c^{(\ell)} = \begin{pmatrix} \alpha_\ell \mathbf{I}_K + \beta_\ell (\mathbf{1}_K \mathbf{1}_K^\top - \mathbf{I}_K) & \rho_\ell \mathbf{1}_K \\ \varsigma_\ell \mathbf{1}_K^\top & \mu_\ell \end{pmatrix}. \tag{47}$$

We next remove the row-shift directions invisible to softmax. Since every content representation is one-hot, $\mathbf{Z}\mathbf{1}_D = \mathbf{1}_T$, and adding $\mathbf{w}\mathbf{1}_D^\top$ for $\mathbf{w} \in \mathbb{R}^D$ changes the pre-softmax score only by a row-dependent constant in either layer:

$$\mathbf{Z}(\mathbf{W}_c^{(\ell)} + \mathbf{w}\mathbf{1}_D^\top)\mathbf{Z}^\top = \mathbf{Z}\mathbf{W}_c^{(\ell)}\mathbf{Z}^\top + (\mathbf{Z}\mathbf{w})\mathbf{1}_T^\top, \tag{48}$$

$$\mathbf{S}^{(1)}\mathbf{Z}(\mathbf{W}_c^{(\ell)} + \mathbf{w}\mathbf{1}_D^\top)\mathbf{Z}^\top = \mathbf{S}^{(1)}\mathbf{Z}\mathbf{W}_c^{(\ell)}\mathbf{Z}^\top + (\mathbf{S}^{(1)}\mathbf{Z}\mathbf{w})\mathbf{1}_T^\top. \tag{49}$$

Therefore $\mathbf{W}_c^{(\ell)}$ and $\mathbf{W}_c^{(\ell)} + \mathbf{w}\mathbf{1}_D^\top$ induce the same attention maps. Compatibility with vertex relabeling requires $C_\tau \mathbf{w} = \mathbf{w}$ for every $\tau \in S_K$, so $\mathbf{w}$ has one common node coordinate and one independent mask coordinate, namely $\mathbf{w} = (w_{\text{node}}\mathbf{1}_K^\top, w_{\text{mask}})^\top$. The unique row-centered representative within each such equivalence class is

$$\overline{\mathbf{W}}_c^{(\ell)} = \mathbf{W}_c^{(\ell)} \left(\mathbf{I}_D - \frac{1}{D} \mathbf{1}_D \mathbf{1}_D^\top \right), \qquad \overline{\mathbf{W}}_c^{(\ell)} \mathbf{1}_D = \mathbf{0}. \tag{50}$$

For a row-centered content block in the same fixed form, the coefficients satisfy

$$\alpha_\ell + (K-1)\beta_\ell + \rho_\ell = 0, \qquad K\varsigma_\ell + \mu_\ell = 0. \tag{51}$$

These two independent constraints leave three coordinates from the five-dimensional content fixed space.

For the positional block, the relevant distinction is between edges sharing a slot and edges occupying different slots, not between named arms. Write

$$\mathbf{W}_p^{(\ell)} = \begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{R} & \mathbf{D}^{(\ell)} \end{pmatrix}. \tag{52}$$

Let P_σ be the permutation matrix on the m edge slots, so $Q_\sigma = (P_\sigma \otimes \mathbf{I}_2) \oplus \mathbf{I}_4$. The condition $Q_\sigma \mathbf{W}_p^{(\ell)} Q_\sigma^\top = \mathbf{W}_p^{(\ell)}$ gives

$$(P_\sigma \otimes \mathbf{I}_2)\mathbf{X}(P_\sigma^\top \otimes \mathbf{I}_2) = \mathbf{X}, \qquad (P_\sigma \otimes \mathbf{I}_2)\mathbf{Y} = \mathbf{Y}, \qquad \mathbf{R}(P_\sigma^\top \otimes \mathbf{I}_2) = \mathbf{R}. \tag{53}$$

Partition $\mathbf{X}$ into blocks $\mathbf{X}_{rr'}$ indexed by ordered edge pairs. The first identity gives $\mathbf{X}_{rr'} = \mathbf{X}_{\sigma(r)\sigma(r')}$, and Proposition A.2 identifies exactly the same-edge and different-edge orbits. Since the source and destination roles are preserved, there exist independent 2×2 matrices $\mathbf{A}_=^{(\ell)}$ and $\mathbf{A}_\neq^{(\ell)}$ such that

$$\mathbf{X}_{rr} = \mathbf{A}_=^{(\ell)}, \qquad \mathbf{X}_{rr'} = \mathbf{A}_\neq^{(\ell)} \quad (r \neq r'), \qquad \mathbf{X} = \mathbf{I}_m \otimes \mathbf{A}_=^{(\ell)} + (\mathbf{1}_m \mathbf{1}_m^\top - \mathbf{I}_m) \otimes \mathbf{A}_\neq^{(\ell)}. \tag{54}$$

The remaining fixed-point equations force the edge-to-special and special-to-edge blocks to be constant across the edge index, while the special-to-special block is unrestricted:

$$\mathbf{Y} = \mathbf{1}_m \otimes \mathbf{B}^{(\ell)}, \qquad \mathbf{R} = \mathbf{1}_m^\top \otimes \mathbf{C}^{(\ell)}. \tag{55}$$

Combining the blocks gives

$$\mathbf{W}_p^{(\ell)} = \begin{pmatrix} \mathbf{I}_m \otimes \mathbf{A}_{=}^{(\ell)} + (\mathbf{1}_m\mathbf{1}_m^\top - \mathbf{I}_m) \otimes \mathbf{A}_{\neq}^{(\ell)} & \mathbf{1}_m \otimes \mathbf{B}^{(\ell)} \\ \mathbf{1}_m^\top \otimes \mathbf{C}^{(\ell)} & \mathbf{D}^{(\ell)} \end{pmatrix}. \tag{56}$$

As with the content block, adding $\mathbf{w}\mathbf{1}_T^\top$ with $\mathbf{w} \in \mathbb{R}^T$ to a positional block changes the corresponding pre-softmax score only by a row-dependent constant:

$$\mathbf{Z}\mathbf{W}_c^{(1)}\mathbf{Z}^\top + \mathbf{W}_p^{(1)} + \mathbf{w}\mathbf{1}_T^\top = \left(\mathbf{Z}\mathbf{W}_c^{(1)}\mathbf{Z}^\top + \mathbf{W}_p^{(1)}\right) + \mathbf{w}\mathbf{1}_T^\top, \tag{57}$$

$$\mathbf{S}^{(1)}\left(\mathbf{Z}\mathbf{W}_c^{(2)}\mathbf{Z}^\top + \mathbf{W}_p^{(2)} + \mathbf{w}\mathbf{1}_T^\top\right) = \mathbf{S}^{(1)}\left(\mathbf{Z}\mathbf{W}_c^{(2)}\mathbf{Z}^\top + \mathbf{W}_p^{(2)}\right) + (\mathbf{S}^{(1)}\mathbf{w})\mathbf{1}_T^\top. \tag{58}$$

Hence $\mathbf{W}_p^{(\ell)}$ and $\mathbf{W}_p^{(\ell)} + \mathbf{w}\mathbf{1}_T^\top$ induce the same attention maps. Compatibility with edge serialization requires $Q_\sigma \mathbf{w} = \mathbf{w}$ for every $\sigma \in S_m$, so $\mathbf{w}$ is constant on source positions, constant on destination positions, and unrestricted on the four special positions:

$$\mathbf{w} = w_S \sum_{r=1}^{m} \mathbf{e}_{s_r} + w_D \sum_{r=1}^{m} \mathbf{e}_{d_r} + w_{q_s}\mathbf{e}_{q_s} + w_{q_t}\mathbf{e}_{q_t} + w_{a_1}\mathbf{e}_{a_1} + w_{a_2}\mathbf{e}_{a_2}. \tag{59}$$

The unique row-centered representative within each row-shift equivalence class is

$$\overline{\mathbf{W}}_p^{(\ell)} = \mathbf{W}_p^{(\ell)}\left(\mathbf{I}_T - \frac{1}{T}\mathbf{1}_T\mathbf{1}_T^\top\right), \qquad \overline{\mathbf{W}}_p^{(\ell)}\mathbf{1}_T = \mathbf{0}. \tag{60}$$

For a row-centered positional block in the same fixed form, the coefficients satisfy

$$\left(\mathbf{A}_{=}^{(\ell)} + (m-1)\mathbf{A}_{\neq}^{(\ell)}\right)\mathbf{1}_2 + \mathbf{B}^{(\ell)}\mathbf{1}_4 = \mathbf{0}_2, \qquad m\mathbf{C}^{(\ell)}\mathbf{1}_2 + \mathbf{D}^{(\ell)}\mathbf{1}_4 = \mathbf{0}_4. \tag{61}$$

Both quotient dimensions are independent of R and L.

It remains to show that centering describes the actual trajectory, rather than merely an equivalent parameterization. For one masked example, let $\mathbf{\Gamma}_\ell$ denote the total derivative of its masked loss with respect to the layer-ℓ pre-softmax logits. Since the row-softmax Jacobian annihilates the constant key direction,

$$\mathbf{\Gamma}_\ell\mathbf{1}_T = \mathbf{0}, \qquad \ell \in \{1, 2\}. \tag{62}$$

With expectations over the population and mask law, the parameter gradients satisfy

$$\nabla_{\mathbf{W}_p^{(1)}}\mathcal{L} = \mathbb{E}[\mathbf{\Gamma}_1], \qquad \nabla_{\mathbf{W}_c^{(1)}}\mathcal{L} = \mathbb{E}[\mathbf{Z}^\top\mathbf{\Gamma}_1\mathbf{Z}], \tag{63}$$

$$\nabla_{\mathbf{W}_p^{(2)}}\mathcal{L} = \mathbb{E}[\mathbf{S}^{(1)\top}\mathbf{\Gamma}_2], \qquad \nabla_{\mathbf{W}_c^{(2)}}\mathcal{L} = \mathbb{E}[\mathbf{Z}^\top\mathbf{S}^{(1)\top}\mathbf{\Gamma}_2\mathbf{Z}]. \tag{64}$$

Since $\mathbf{Z}\mathbf{1}_D = \mathbf{1}_T$, it follows that

$$\nabla_{\mathbf{W}_p^{(1)}}\mathcal{L}\,\mathbf{1}_T = \mathbf{0}, \qquad \nabla_{\mathbf{W}_p^{(2)}}\mathcal{L}\,\mathbf{1}_T = \mathbf{0}, \tag{65}$$

$$\nabla_{\mathbf{W}_c^{(1)}}\mathcal{L}\,\mathbf{1}_D = \mathbb{E}[\mathbf{Z}^\top\mathbf{\Gamma}_1\mathbf{1}_T] = \mathbf{0}, \qquad \nabla_{\mathbf{W}_c^{(2)}}\mathcal{L}\,\mathbf{1}_D = \mathbb{E}[\mathbf{Z}^\top\mathbf{S}^{(1)\top}\mathbf{\Gamma}_2\mathbf{1}_T] = \mathbf{0}. \tag{66}$$

Hence, along population gradient flow,

$$\frac{d}{dt}\left(\mathbf{W}_c^{(\ell)}(t)\mathbf{1}_D\right) = \mathbf{0}, \qquad \frac{d}{dt}\left(\mathbf{W}_p^{(\ell)}(t)\mathbf{1}_T\right) = \mathbf{0}. \tag{67}$$

Since $\mathbf{W}_c^{(\ell)}(0) = \mathbf{0}$ and $\mathbf{W}_p^{(\ell)}(0) = \mathbf{0}$, we obtain

$$\mathbf{W}_c^{(\ell)}(t)\mathbf{1}_D = \mathbf{0}, \qquad \mathbf{W}_p^{(\ell)}(t)\mathbf{1}_T = \mathbf{0} \tag{68}$$

for every $t \geq 0$. Thus centering holds on the trajectory itself, completing the fixed-space and row-shift reduction without imposing a circuit-specific parameter restriction. □

A.4 Layer-2 Pair Blindness

Write

$$\mathcal{V}_p := \{\mathbf{W} \in \mathbb{R}^{T\times T} : Q_\sigma \mathbf{W} Q_\sigma^\top = \mathbf{W} \ \ \forall \sigma \in S_m, \ \ \mathbf{W}\mathbf{1}_T = \mathbf{0}\},$$
$$\mathbf{P}_{\text{role}} := \Big(\frac{\mathbf{1}_m\mathbf{1}_m^\top}{m} \otimes \mathbf{I}_2\Big) \oplus \mathbf{I}_4, \qquad \mathcal{V}_2 := \{\mathbf{W} \in \mathcal{V}_p : \mathbf{A}_= = \mathbf{A}_\neq\}. \tag{69}$$

By Lemma 2.1, both positional blocks lie in $\mathcal{V}_p$ along the zero-initialized population flow. For $\mathbf{W} \in \mathcal{V}_p$, membership in $\mathcal{V}_2$ is equivalent to $\mathbf{P}_{\text{role}}\mathbf{W} = \mathbf{W}$, and also to all graph-source columns of $\mathbf{W}$ being identical and all graph-destination columns being identical. Within this subspace, a Layer-2 positional score therefore depends on a graph key only through its source or destination role.

Lemma A.5 (Layer-2 pair blindness)**.** *Under the zero-initialized symmetric population gradient flow, $\mathbf{W}_p^{(2)}(t) \in \mathcal{V}_2$ for all $t \geq 0$. The restriction holds separately for each mask-conditioned gradient rather than through cancellation between mask cases.*

Proof. Let $\mathbf{P}_{\text{role}}$ be as in equation 69. For a supervised query $i \in M$, its original token is masked. Content symmetry makes all node-key content scores equal, and positional symmetry makes them equal within each graph role. Consequently, $S^{(1)}_{i,s_r} = S^{(1)}_{i,s_{r'}}$ and $S^{(1)}_{i,d_r} = S^{(1)}_{i,d_{r'}}$, so $\mathbf{P}_{\text{role}}\mathbf{S}^{(1)\top}\mathbf{e}_i = \mathbf{S}^{(1)\top}\mathbf{e}_i$. For the fixed masked input $\mathbf{x}_0[M]$, let $\ell_{\text{ex}} = \ell_{\text{ex}}(\theta; \mathbf{x}_0, M)$ be its masked loss and let $\mathbf{\Gamma}_2$ denote its total derivative with respect to the Layer-2 logits; its nonzero rows are confined to M. Differentiating the exact forward pass gives

$$\mathbf{P}_{\text{role}}\nabla_{\mathbf{W}_p^{(2)}}\ell_{\text{ex}} = \sum_{i\in M}(\mathbf{P}_{\text{role}}\mathbf{S}^{(1)\top}\mathbf{e}_i)\mathbf{\Gamma}_{2,i:} \tag{70}$$
$$= \sum_{i\in M}(\mathbf{S}^{(1)\top}\mathbf{e}_i)\mathbf{\Gamma}_{2,i:} = \nabla_{\mathbf{W}_p^{(2)}}\ell_{\text{ex}}. \tag{71}$$

Population averaging and zero initialization imply

$$\mathbf{W}_p^{(2)}(t) = \int_0^t \mathbf{P}_{\text{role}}\dot{\mathbf{W}}_p^{(2)}(s)\,ds = \mathbf{P}_{\text{role}}\mathbf{W}_p^{(2)}(t). \tag{72}$$

For each source/destination block, this equality requires

$$\big(\mathbf{A}_=^{(2)}\big)_{\nu,\nu'} = \frac{\big(\mathbf{A}_=^{(2)}\big)_{\nu,\nu'} + (m-1)\big(\mathbf{A}_\neq^{(2)}\big)_{\nu,\nu'}}{m}, \qquad \nu,\nu' \in \{1,2\}. \tag{73}$$

Thus $\mathbf{A}_=^{(2)} = \mathbf{A}_\neq^{(2)}$, proving the claim. The two block indices follow the source–destination order of $\mathcal{I}_r = (s_r, d_r)$. □

Remark. Layer-2 position alone cannot select a destination within its pool, so matching must use content. The same argument applies with frozen symmetry-fixed content in Stage II of Section 3.2, and to other target placements with the same masking pattern.

B PROOFS FOR SECTION 3

Evaluated states have symmetry-fixed content, $W_p^{(1)} \in \mathcal{V}_p$, and $W_p^{(2)} \in \mathcal{V}_2$. The joint flow starts at zero; staged learning follows Section 3.2. Write $y_1 = v_{L-2}$, $y_2 = v^\star = v_{L-1}$, $s^\star = s_{r_\star}$, and $d^\star = d_{r_\star}$. The constant C_1 is the one in equation 11.

In each calculation, the masked input is specified first; forward quantities then refer to that input without a mask-state subscript. When several conditional problems are summed or compared, $\mathcal{L}(e;t)$ and $R(e;t) = \mathcal{L}(e;t) - \ell_{\min}$ denote the conditional population loss and excess risk for problem $e \in \mathcal{E}_4$. We omit t when the evaluated parameter state is fixed. The joint objective remains $\mathcal{L}(\theta)$, without a conditional-problem argument. For the both-masked input, define

$$\varepsilon_i(\theta; \mathbf{x}_0) = 1 - H^{(2)}_{a_i, y_i}, \qquad i \in \{1, 2\}. \tag{74}$$

For a probability vector $\mathbf{h}$ on the node and mask tokens and a node target y, define $F_y(\mathbf{h}) = \log(\sum_{v \in \mathcal{V}} e^{\alpha h_v}) - \alpha h_y$.

Lemma B.1 (Loss floor, output-value purity, and greedy error)**.** *We have*

$$C_1(1 - h_y) \le F_y(\mathbf{h}) - \ell_{\min} \le 2\alpha(1 - h_y). \tag{75}$$

In particular, $\mathcal{R} \ge 0$. Any greedy node prediction $\widehat{y} \in \arg\max_{v \in \mathcal{V}} h_v$ satisfies $\mathbf{1}\{\widehat{y} \ne y\} \le 2(1-h_y)$ under every tie-breaking rule, and $h_y > 1/2$ makes y the unique maximizer. Fix any conditional problem $e \in \mathcal{E}_4$, let $i \in \{a_1, a_2\}$ be its supervised position and $y = [\mathbf{x}_0]_i$ its target, and evaluate the forward pass in that problem. Every population input satisfies

$$C_1\big(1 - H^{(2)}_{i,y}\big) \le \mathcal{L}(e) - \ell_{\min} \le 4\mathcal{R}. \tag{76}$$

On the both-masked input, the two output impurities also satisfy

$$C_1(\varepsilon_1 + \varepsilon_2) \le R(1 \mid \varnothing) + R(2 \mid \varnothing) \le 4\mathcal{R}. \tag{77}$$

Proof. The rows of $\mathbf{H}^{(2)}$ are probability vectors because both attention matrices are row-stochastic and the original values are one-hot. Since $F_y(\mathbf{e}_y) = \ell_{\min}$, convexity gives

$$F_y(\mathbf{h}) - F_y(\mathbf{e}_y) \ge \langle \nabla F_y(\mathbf{e}_y), \mathbf{h} - \mathbf{e}_y \rangle \tag{78}$$

$$= \frac{\alpha(K-1)}{e^\alpha + K - 1}(1 - h_y) + \frac{\alpha}{e^\alpha + K - 1} \sum_{v \in \mathcal{V} \setminus \{y\}} h_v \tag{79}$$

$$\ge C_1(1 - h_y). \tag{80}$$

The gradient on the node coordinates has infinity norm at most α, and its mask coordinate is zero. Thus

$$|F_y(\mathbf{h}) - F_y(\mathbf{e}_y)| \le \alpha \|\mathbf{h} - \mathbf{e}_y\|_1 = 2\alpha(1 - h_y). \tag{81}$$

If $\widehat{y} \ne y$, then $h_{\widehat{y}} \ge h_y$ and $h_{\widehat{y}} + h_y \le 1$, so $h_y \le 1/2$ and $2(1 - h_y) \ge 1$. If $h_y > 1/2$, every other coordinate is at most $1 - h_y < h_y$. Nonnegativity of every elementary excess gives $\mathcal{R} \ge 0$.

Each elementary loss is constant on the clean-sequence orbit of Proposition A.3, so its value on any input equals the population subloss. Applying the scalar bounds to each output row and using nonnegativity of the other conditional excesses proves the population bounds. □

B.1 EXACT GRADIENTS AND PROJECTIONS

Fix an elementary prediction problem, its masked input $\widetilde{\mathbf{x}}$, supervised position $i \in \{a_1, a_2\}$, and target $y = [\mathbf{x}_0]_i$. Let $\ell_{\text{pred}} = F_y(\mathbf{h})$ be its elementary loss. We suppress the clean-sequence argument and use column vectors for attention rows.

Forward quantities. Define the layer-wise score, attention, and hidden-state quantities by

$$\mathbf{\Lambda}^{(1)} = \mathbf{Z} W_c^{(1)} \mathbf{Z}^\top + W_p^{(1)}, \qquad \mathbf{S}^{(1)} = \operatorname{softmax}(\mathbf{\Lambda}^{(1)}), \qquad \mathbf{H}^{(1)} = \mathbf{S}^{(1)} \mathbf{Z}, \tag{82}$$

$$\mathbf{\Omega} = \mathbf{Z} W_c^{(2)} \mathbf{Z}^\top + W_p^{(2)}, \qquad \mathbf{\Lambda}^{(2)} = \mathbf{S}^{(1)} \mathbf{\Omega}. \tag{83}$$

Prediction quantities. For the supervised row, write

$$\mathbf{s} := (S^{(2)}_{i,:})^\top, \qquad \mathbf{h} := \mathbf{H}^{(1)\top}\mathbf{s}, \qquad \mathbf{p} := (\Pi_{i,:})^\top = \mathrm{softmax}(\alpha \mathbf{h}[\mathcal{V}]), \tag{84}$$

and for each Layer-1 row k, let $\mathbf{q}_k := (S^{(1)}_{k,:})^\top$.

Backpropagation quantities. For any probability vector $\mathbf{u}$, define the softmax Jacobian

$$J(\mathbf{u}) := \mathrm{diag}(\mathbf{u}) - \mathbf{u}\mathbf{u}^\top. \tag{85}$$

Then set

$$\mathbf{g}^{\mathrm{out}} := \alpha \begin{pmatrix} \mathbf{p} - \mathbf{e}_y \\ 0 \end{pmatrix}, \qquad \boldsymbol{\xi} := J(\mathbf{s})\mathbf{H}^{(1)}\mathbf{g}^{\mathrm{out}}. \tag{86}$$

Here $\mathbf{h}[\mathcal{V}]$ denotes the restriction of $\mathbf{h}$ to the node coordinates, and the final coordinate of $\mathbf{g}^{\mathrm{out}}$ is the mask coordinate.

The matrix $\boldsymbol{\Omega}$ is the Layer-2 score matrix before left multiplication by $\mathbf{S}^{(1)}$; K continues to denote the node-vocabulary size. Direct differentiation gives the full positional gradients

$$\mathbf{G}_2 : = \nabla_{W_p^{(2)}} \ell_{\mathrm{pred}} = \mathbf{q}_i \boldsymbol{\xi}^\top, \qquad (\mathbf{G}_1)^\top_{k,:} = J(\mathbf{q}_k)\left[S^{(2)}_{i,k}\mathbf{Z}\mathbf{g}^{\mathrm{out}} + \mathbf{1}_{\{k=i\}}\boldsymbol{\Omega}\boldsymbol{\xi}\right], \tag{87}$$

$$\nabla_{W_c^{(1)}} \ell_{\mathrm{pred}} = \mathbf{Z}^\top \mathbf{G}_1 \mathbf{Z}, \qquad \nabla_{W_c^{(2)}} \ell_{\mathrm{pred}} = \mathbf{Z}^\top \mathbf{G}_2 \mathbf{Z}, \tag{88}$$

where $\mathbf{G}_1 = \nabla_{W_p^{(1)}} \ell_{\mathrm{pred}}$. The second term in $\mathbf{G}_1$ is the query-backpropagation path through the second layer. It cannot be discarded for the supervised row, even though it vanishes on graph rows.

Write $\mathbf{C}_n = \mathbf{I}_n - n^{-1}\mathbf{1}_n\mathbf{1}_n^\top$, extending the centering notation $\mathbf{C}_D$ of Section 3.2, and let $\mathcal{I}_S = \{s_r\}_{r\in[m]}$ and $\mathcal{I}_D = \{d_r\}_{r\in[m]}$ denote the graph source and destination positions. For $\mathcal{J} \subseteq [T]$, write $\mathbf{1}_{\mathcal{J}} = \sum_{j\in\mathcal{J}} \mathbf{e}_j$. The output-probability matrix remains $\boldsymbol{\Pi}$. For a conditional problem e, write $g(e) = \nabla_\theta \mathcal{L}(e)$ for its population gradient. It is obtained by averaging the full gradients over clean sequences and projecting onto the centered fixed spaces of Lemmas 2.1 and A.5. For the positional blocks, the Frobenius-orthogonal projections are

$$\begin{aligned} \mathrm{Proj}_{\mathcal{V}_p}\, \mathbf{G} &= \Big(\frac{1}{m!}\sum_{\sigma\in S_m} Q_\sigma \mathbf{G} Q_\sigma^\top\Big)\mathbf{C}_T, \\ \mathrm{Proj}_{\mathcal{V}_2}\, \mathbf{G} &= \mathbf{P}_{\mathrm{role}}\mathbf{G}\,\mathbf{P}_{\mathrm{role}}\mathbf{C}_T. \end{aligned} \tag{89}$$

The group average requires no enumeration: it replaces every same-edge block and every different-edge block by

$$\mathbf{G}_= = \frac{1}{m}\sum_r \mathbf{G}[\mathcal{I}_r, \mathcal{I}_r], \qquad \mathbf{G}_{\neq} = \frac{1}{m(m-1)}\sum_{r\neq r'} \mathbf{G}[\mathcal{I}_r, \mathcal{I}_{r'}], \tag{90}$$

respectively, replaces the edge-to-special and special-to-edge blocks by their edge averages, and keeps $\mathbf{G}[\mathcal{I}_{\mathrm{sp}}, \mathcal{I}_{\mathrm{sp}}]$. The content projection averages the node-diagonal, node-off-diagonal, node-to-mask, mask-to-node, and mask-to-mask entry classes and right-multiplies by $\mathbf{C}_D$. At an invariant parameter state these projections leave the population gradients unchanged.

Mask-resolved loss rates. The population flow is $\dot\theta = -\bar g$ with $\bar g = \frac{1}{4}\sum_{f\in\mathcal{E}_4} g(f)$. The chain rule gives, for every $e \in \mathcal{E}_4$,

$$\dot{\mathcal{L}}(e;t) = -\langle g(e;t), \bar g(t)\rangle_F = -\frac{1}{4}\sum_{f\in\mathcal{E}_4}\langle g(e;t), g(f;t)\rangle_F, \tag{91}$$

where $\langle\cdot,\cdot\rangle_F$ sums the Frobenius inner products over the four parameter blocks. The Gram matrix with entries $\langle g(e), g(f)\rangle_F$, indexed by $e, f \in \mathcal{E}_4$, is positive semidefinite, but its off-diagonal entries have no definite sign: through the shared parameters one conditional problem can help or interfere with another, and an individual conditional loss need not decrease monotonically.

B.2 PROOF OF THE LOCAL ENDPOINT-FIRST SEPARATION

The joint-training calculation uses $q = \exp(\alpha/T) > 1$. Unprimed probabilities below use the both-masked count profile; primed probabilities use the single-masked count profile. A prime labels the second profile, whereas time derivatives are denoted by dots:

$$p_j := \frac{q^j}{q^{R+1} + (K-R)q^2 + (R-1)q}, \qquad \Delta_{\mathrm{root}} := p_{R+1} - p_2, \tag{92}$$

$$p'_j := \frac{q^j}{q^{R+1} + q^3 + (K-R-1)q^2 + (R-1)q}, \qquad \Delta_{\mathrm{reveal}} := p'_3 - p'_2. \tag{93}$$

Define

$$\Xi_{R,L,\alpha} := \frac{m-1}{2m} + \frac{R}{2m}(\Delta_{\mathrm{root}} + p'_{R+1} - p'_2) - \frac{m-1}{4m}\Delta_{\mathrm{reveal}} > 0. \tag{94}$$

The exact initial-rate identities are

$$\dot{\mathcal{L}}(1 \mid \varnothing; 0) - \dot{\mathcal{L}}(2 \mid \varnothing; 0) = \frac{\alpha^2}{T^3}\Xi_{R,L,\alpha} > 0, \tag{95}$$

$$\dot{\mathcal{L}}(1 \mid 2; 0) - \dot{\mathcal{L}}(2 \mid 1; 0) = (1 + \Delta_{\mathrm{reveal}})\frac{\alpha^2}{T^3}\Xi_{R,L,\alpha} > 0. \tag{96}$$

Proof of Theorem 3.1. At zero initialization both attention matrices equal $T^{-1}\mathbf{1}_T\mathbf{1}_T^\top$. Fix any of the four conditional problems and evaluate all forward quantities in that problem. All first-layer values equal the input histogram $\mathbf{f} = T^{-1}\mathbf{Z}^\top\mathbf{1}_T$, so $\boldsymbol{\xi} = 0$ and both second-layer parameter gradients vanish. For $\mathbf{z}_j = \mathbf{e}_{\widetilde{x}_j}$, put $d_j^{\mathrm{out}} := (\mathbf{g}^{\mathrm{out}})^\top(\mathbf{z}_j - \mathbf{f})$. Then

$$(G_1)_{k,j}(0) = T^{-2}d_j^{\mathrm{out}}. \tag{97}$$

Thus all rows of the positional gradient are identical. Edge averaging makes them constant within the source and destination pools, and the first-layer pair-incidence projection is zero.

In the both-masked input the root has multiplicity $R+1$, the $K-R$ internal or target-endpoint nodes have multiplicity 2, and the $R-1$ non-target endpoints have multiplicity 1; there are two masks. Revealing either answer increases one multiplicity from 2 to 3 and leaves one mask. Every supervised target still has multiplicity 2. These counts give p_j, p'_j and equality of the paired initial losses. They also imply that the population first-content gradients agree between $1 \mid \varnothing$ and $2 \mid \varnothing$, and between $1 \mid 2$ and $2 \mid 1$: after relabeling, the count profile and the distinguished target count are the same. Hence those content gradients cancel when taking the two target-rate differences.

For the both-masked pair, the first-positional block of $g(2 \mid \varnothing) - g(1 \mid \varnothing)$ has the common row

$$\frac{\alpha}{T^2}\left(\frac{1}{m}\mathbf{1}_{\mathcal{I}_S} - \mathbf{e}_{q_t}\right)^\top. \tag{98}$$

Let $\bar{G}$ be the first-positional block of $\bar{g}$ at zero, and let $\bar{G}_{k,S}$ denote its value in any source column. Direct averaging of the four task contributions gives

$$\bar{G}_{k,S} - \bar{G}_{k,q_t} = \frac{\alpha}{T^2}\Xi_{R,L,\alpha}. \tag{99}$$

The source-minus-query probability differences are $R\Delta_{\mathrm{root}}/m$ in each both-masked case, $R(p'_{R+1} - p'_2)/m - \Delta_{\mathrm{reveal}}$ in $1 \mid 2$, and $[R(p'_{R+1} - p'_2) + \Delta_{\mathrm{reveal}}]/m$ in $2 \mid 1$. The target indicators contribute $(m-1)/(2m)$ after averaging; histogram-centering terms cancel. Taking the Frobenius inner product with the common-row contrast proves the first initial-rate identity.

For the single-masked pair, the corresponding common-row difference is

$$\frac{\alpha}{T^2}\left[(1 + \Delta_{\mathrm{reveal}})\left(\frac{1}{m}\mathbf{1}_{\mathcal{I}_S} - \mathbf{e}_{q_t}\right) + p'_3(\mathbf{e}_{a_1} - \mathbf{e}_{a_2})\right]^\top. \tag{100}$$

The average gradient has identical a_1 and a_2 columns at zero, so the last term has zero inner product with $\bar{G}$. This proves the second identity.

Since $p'_{R+1} \geq p'_3$, we have $0 < \Delta_{\text{reveal}} < p'_3 < 1/2$. Both root-to-count-two gaps are positive, so

$$\Xi_{R,L,\alpha} \geq \frac{m-1}{2m} - \frac{m-1}{4m}\Delta_{\text{reveal}} > \frac{3(m-1)}{8m} > 0. \tag{101}$$

Smoothness of the finite-population gradient field gives the first-order expansions of the paired loss differences.

For the finite-threshold assertion, let $R_0 = R_E(0) = R_P(0)$ and $\Delta_0 = \dot{R}_P(0) - \dot{R}_E(0) > 0$. Since $R_E + R_P = 4(\mathcal{L} - \ell_{\min})$, the joint gradient flow gives

$$c_E := -\dot{R}_E(0) = 2\|\nabla\mathcal{L}(0)\|_F^2 + \frac{\Delta_0}{2} > 0. \tag{102}$$

By continuity, choose $\tau_E^{\text{loc}} > 0$ such that $-\dot{R}_E \geq c_E/2$ and $\dot{R}_P - \dot{R}_E \geq \Delta_0/2$ on $[0, \tau_E^{\text{loc}}]$. Set

$$h_E := \min\{R_0/2, c_E\tau_E^{\text{loc}}/2\}, \qquad \eta := R_0 - h, \qquad 0 < h < h_E. \tag{103}$$

Integration gives $R_E(t) \leq R_0 - c_E t/2$, so $0 < t_E \leq 2h/c_E < \tau_E^{\text{loc}}$. For $0 < t \leq t_E$, $R_P(t) - R_E(t) \geq \Delta_0 t/2 > 0$. Before t_E, $R_E(t) > \eta$, so R_P has not previously crossed η either. At t_E, continuity gives $R_E(t_E) = \eta < R_P(t_E)$, and hence $t_E < \tau_P(\eta)$. The threshold lies near R_0; no near-zero risk is needed for this local assertion. □

This local joint-training asymmetry motivates Stage I; it does not establish endpoint readiness or decoding success.

B.3 Stage I: endpoint provenance, carriers, and query transfer

Proof of Corollary 3.2. Fix either endpoint problem, $2 \mid \varnothing$ or $2 \mid 1$, and evaluate the forward pass in that problem. Each conditional excess is nonnegative, so Lemma B.1 gives

$$C_1\big(1 - H^{(2)}_{a_2,v^\star}\big) \leq R_E \leq \eta. \tag{104}$$

Divide by C_1. The inequality applies at any evaluated state; the prescribed saturated entry state of the staged model can use a sharper impurity bound ε_E than η/C_1. □

For the statewise endpoint arguments, use the both-masked input. Following Lemma 2.1, $W^{(1)}_{c,ii} - W^{(1)}_{c,ij} = \alpha_1 - \beta_1$ for distinct node identities. We assume $\alpha_1 - \beta_1 \leq 0$ for the endpoint occurrence bound and use the selected content, $\alpha_1 - \beta_1 = -a$, for its selected-content consequence. Set

$$B_{\text{self}} := W^{(1)}_{p,d^\star,d^\star} - W^{(1)}_{p,d^\star,s^\star}. \tag{105}$$

Lemma B.2 (Endpoint provenance under selected content)**.** *The endpoint occurrence and selected-content bounds are*

$$Q_{a_2,d^\star} \leq 2\varepsilon_2 + S^{(2)}_{a_2,d^\star} S^{(1)}_{d^\star,d^\star}, \tag{106}$$

$$1 - Q_{a_2,q_t} \leq 3\varepsilon_2 + e^{B_{\text{self}} - a}. \tag{107}$$

In particular, $\varepsilon_2 \leq \varepsilon_E$ and $a \geq B_{\text{self}} + \log(1/\varepsilon_E)$ imply $Q_{a_2,q_t} \geq 1 - 4\varepsilon_E$.

Proof. Put $n = L - 1$ and retain the arm labels $h \in [R]$ and target arm $h_\star$ from Section 2. Write s^h_j, d^h_j for the serialized positions of edge $(v_{h,j-1}, v_{h,j})$. Thus $v^\star = v_n$, $s^\star = s^{h_\star}_n$, and $d^\star = d^{h_\star}_n$. The endpoint token occurs only at q_t and $d^\star$, so

$$1 - \varepsilon_2 = Q_{a_2,q_t} + Q_{a_2,d^\star}. \tag{108}$$

We bound the latter term by splitting its intermediate rows. If a graph row k is outside $\{s^\star, d^\star\}$, exclude the target edge, the edge of k, and the at most one edge whose destination equals its query token. Since $m \geq 4$, a remaining destination is a wrong-token key with the same unpaired positional category and the same content score as $d^\star$. For $k \in \{q_s, a_1, a_2\}$, another destination likewise has

the same score as $d^\star$. For $k = q_t$, another destination has at least its score because $\alpha_1 - \beta_1 \le 0$. Therefore

$$S^{(1)}_{k,d^\star} \le 1 - H^{(1)}_{k,v^\star} \qquad (k \notin \{s^\star, d^\star\}), \tag{109}$$

$$\sum_{k \notin \{s^\star, d^\star\}} S^{(2)}_{a_2,k} S^{(1)}_{k,d^\star} \le \varepsilon_2. \tag{110}$$

For any non-target arm $h \neq h_\star$, the rows $s_n^{h_\star}$ and s_n^h have equal Layer-1 softmax denominators. Indeed, each query has its only equal source at its paired source position, its only equal destination at one unpaired destination, different tokens at both special node keys, and mask tokens at both answer positions. All remaining graph keys are different from the query token, with identical counts in each positional category. Their paired-destination numerators are equal as well. The masked a_2 query gives all internal-node tokens equal mass in $H^{(1)}_{a_2,:}$, and Layer-2 pair blindness then makes its attention constant within the internal-source class. Consequently,

$$S^{(2)}_{a_2,s_n^{h_\star}} S^{(1)}_{s_n^{h_\star},d_n^{h_\star}} = S^{(2)}_{a_2,s_n^h} S^{(1)}_{s_n^h,d_n^h} \tag{111}$$

$$\le S^{(2)}_{a_2,s_n^h}(1 - H^{(1)}_{s_n^h,v^\star}) \le \varepsilon_2. \tag{112}$$

The remaining term is $S^{(2)}_{a_2,d^\star} S^{(1)}_{d^\star,d^\star}$, proving the occurrence bound. The chosen content geometry gives

$$\frac{S^{(1)}_{d^\star,d^\star}}{S^{(1)}_{d^\star,s^\star}} = e^{B_{\mathrm{self}} - a}, \qquad S^{(2)}_{a_2,d^\star} S^{(1)}_{d^\star,d^\star} \le e^{B_{\mathrm{self}} - a}. \tag{113}$$

Substitution proves the selected-content bound. The choice of a is made at the evaluated state; no time-uniform positional bound is inferred. □

Proof of Lemma 3.3. Fix the time and the both-masked input, and write $p_k = S^{(2)}_{a_2,k}$, $r_k = S^{(1)}_{k,q_t}$, and $\mathcal{A} = \mathcal{A}_\varepsilon(t; \tilde{\mathbf{x}})$. Row-stochasticity gives $\sum_k p_k = 1$ and $Q_{a_2,q_t} = \sum_k p_k r_k$. For $k \notin \mathcal{A}$, the definition of $\mathcal{A}$ gives $1 - r_k > \varepsilon$. Consequently,

$$\begin{aligned} \varepsilon \left(1 - \sum_{k \in \mathcal{A}} p_k r_k\right) &= \varepsilon \sum_{k \notin \mathcal{A}} p_k + \varepsilon \sum_{k \in \mathcal{A}} p_k (1 - r_k) \\ &\le \sum_{k \notin \mathcal{A}} p_k (1 - r_k) + \sum_{k \in \mathcal{A}} p_k (1 - r_k) \\ &= 1 - Q_{a_2,q_t}. \end{aligned} \tag{114}$$

Division by $\varepsilon > 0$ proves the route-mass bound. Since $0 \le r_k \le 1$, the same lower bound holds for $\sum_{k \in \mathcal{A}} p_k$. The individual retrieval bound follows from the definition of $\mathcal{A}$. If $1 - Q_{a_2,q_t} < \varepsilon$, the lower bound on the route mass is strictly positive, so $\mathcal{A}$ cannot be empty. □

For query transfer, evaluate unprimed quantities on the both-masked input and primed quantities after revealing $a_2 = v^\star$ while keeping a_1 masked, at the same parameters. A prime here labels the revealed input, not a time derivative. Suppose $H^{(2)}_{a_2,v^\star} \ge 1 - \varepsilon_E$ and put $\kappa = S^{(2)}_{a_2,a_1} > 0$.

Lemma B.3 (Coverage transfers endpoint readiness to the predecessor query)**.** *We have*

$$H^{(1)\prime}_{a_1,v^\star} \ge H^{(1)}_{a_1,v^\star} \ge 1 - \frac{\varepsilon_E}{\kappa}. \tag{115}$$

For $b_\star \in (1/2, 1)$, $\varepsilon_E \le \kappa(1 - b_\star)$ gives query endpoint mass at least $b_\star > 1/2$ in both predecessor states. If the logit span of the unprimed Layer-2 a_2 row is M_E, then $\kappa \ge e^{-M_E}/T$.

Proof. On the both-masked input, the network input does not depend on which answer is supervised. Its endpoint impurity satisfies

$$\varepsilon_E \ge \sum_k S^{(2)}_{a_2,k}\big(1 - H^{(1)}_{k,v^\star}\big) \ge \kappa(1 - H^{(1)}_{a_1,v^\star}). \tag{116}$$

Divide by $\kappa > 0$. To compare the predecessor states, reveal $a_2 = v^\star$ while keeping a_1 masked. All other logits in the Layer-1 a_1 row are unchanged. If its previous endpoint numerator and softmax normalizer are N, Z_{norm}, and the old and new positive weights of the a_2 key are w_o, w_n, then

$$H^{(1)\prime}_{a_1,v^\star} - H^{(1)}_{a_1,v^\star} = \frac{N + w_n}{Z_{\text{norm}} - w_o + w_n} - \frac{N}{Z_{\text{norm}}} \tag{117}$$

$$= \frac{w_n(Z_{\text{norm}} - N) + N w_o}{Z_{\text{norm}}(Z_{\text{norm}} - w_o + w_n)} \geq 0. \tag{118}$$

Finally, with $\lambda_k = \Lambda^{(2)}_{a_2,k}$ and $M_E = \max_k \lambda_k - \min_k \lambda_k$, the bound $\exp(\lambda_{a_1})/\sum_k \exp(\lambda_k) \geq e^{-M_E}/T$ proves the softmax coverage claim. The relative-accuracy condition uses the actual κ and impurity at the same state; positivity of κ alone does not imply readiness. □

B.4 Gate identities

Proof of Proposition 3.4. Fix either predecessor state, with both answers masked or only a_1 masked, and evaluate every forward quantity in that state. In $\mathcal{V}_2$, every destination column of $W_p^{(2)}$ is identical. Consequently the effective positional term $\mathbf{S}^{(1)}W_p^{(2)}$ cancels between any two destination columns. With the coefficient notation of Lemma 2.1, the equality contrast is $\alpha_2 - \beta_2 = b > 0$. Content symmetry gives

$$\Lambda^{(2)}_{a_1,d^\star} - \Lambda^{(2)}_{a_1,d_r} = (\alpha_2 - \beta_2)\left(H^{(1)}_{a_1,v^\star} - H^{(1)}_{a_1,[\mathbf{x}_0]_{d_r}}\right) \tag{119}$$

$$= b\left(H^{(1)}_{a_1,v^\star} - H^{(1)}_{a_1,[\mathbf{x}_0]_{d_r}}\right). \tag{120}$$

Normalizing within the destination pool proves the matching identity with equality contrast $\alpha_2 - \beta_2$. The other destination token is distinct from $v^\star$, so its query mass is at most $1 - H^{(1)}_{a_1,v^\star}$. Together with $b > 0$, this proves the lower matching bound.

The terminal node $v^\star$ occurs in no graph-source position. Hence every graph-source content score in row $d^\star$ is the same. Positional symmetry makes the score difference between the paired source and every other source exactly γ_{ds}. Normalization within the source pool proves the second identity and the route certificate. Finally $H^{(2)}_{a_1,y_1} \geq r$, where $r = S^{(2)}_{a_1,d^\star} S^{(1)}_{d^\star,s^\star}$, since the route ends at an original occurrence of y_1. If $r > 1/2$, every wrong output value is strictly smaller. □

B.5 Stage II: predecessor descent and finite-interval attainment

The staged flow. Recall

$$\mathcal{J}_P = \frac{1}{2}\big[\mathcal{L}(1 \mid \varnothing) + \mathcal{L}(1 \mid 2)\big], \qquad R_P = 2(\mathcal{J}_P - \ell_{\min}). \tag{121}$$

Retain the fixed content and fixed trainable positional subspaces $\mathcal{U}_1 \subseteq \mathcal{V}_p, \mathcal{U}_2 \subseteq \mathcal{V}_2$ prescribed in Stage II. Every unfrozen symmetry-allowed coordinate remains trainable. Write $\vartheta = (W_p^{(1)}, W_p^{(2)})$ for the positional state. The flow and its energy identity are

$$\dot{W}_p^{(\ell)} = -\operatorname{Proj}_{\mathcal{U}_\ell} \nabla_{W_p^{(\ell)}} \mathcal{J}_P, \qquad \ell = 1, 2, \tag{122}$$

$$\dot{R}_P = -2\sum_{\ell=1}^{2} \left\| \operatorname{Proj}_{\mathcal{U}_\ell} \nabla_{W_p^{(\ell)}} \mathcal{J}_P \right\|_F^2. \tag{123}$$

Endpoint preservation is prescribed separately from this predecessor objective.

A descent test in the existing positional space. To test the gradient norm, use

$$P_{ds} := \left(\sum_{r=1}^{m} \mathbf{e}_{d_r}\mathbf{e}_{s_r}^\top\right)\mathbf{C}_T, \qquad N_{ds} := \|P_{ds}\|_F^2 = m(1 - 1/T). \tag{124}$$

This symmetry-allowed test direction is not an imposed update rule and requires no sign on γ_{ds}. At the finite entry state $\vartheta_0 = \vartheta(t_0)$, assume $P_{ds} \in \mathcal{U}_1$ and $H^{(1)}_{a_1,v^\star}(t_0) \geq b_\star > b_0 > 1/2$ separately on

each of the two predecessor inputs. With each minimum below taken over those two inputs at the same entry state, set

$$\sigma_{\min} := \min S^{(2)}_{a_1,d^\star}(t_0), \quad \nu_{\min} := \min S^{(1)}_{d^\star,s^\star}(t_0), \quad \omega_{\min} := \min\bigl(1 - S^{(1)}_{d^\star,s^\star}(t_0)\bigr). \tag{125}$$

These constants are positive by finite softmax. Choose the fixed content scale b so that

$$2\alpha(m-1)e^{-b(2b_0-1)} \le \frac{C_1\nu_{\min}\omega_{\min}}{16}. \tag{126}$$

Lemma B.4 (Local strict descent and finite predecessor-risk attainment)**.** *There exist $c_P, \tau > 0$ such that $\dot{R}_P(t) \le -c_P$ on $[t_0, t_0+\tau]$. One may take*

$$\lambda_P := \frac{C_1\sigma_{\min}\nu_{\min}\omega_{\min}}{32}, \qquad c_P := \frac{2\lambda_P^2}{N_{ds}}. \tag{127}$$

Every threshold satisfying $0 < R_P(t_0) - \eta < \min\{R_P(t_0)/2, c_P\tau\}$ has a finite first hitting time, with

$$t_0 < t_P := \inf\{t \ge t_0 : R_P(t) \le \eta\} \le t_0 + \frac{R_P(t_0) - \eta}{c_P} < t_0 + \tau. \tag{128}$$

Proof. Fix an elementary predecessor example in either predecessor state. All forward quantities in the following calculation refer to this fixed input. Write $\sigma_r = S^{(2)}_{a_1,d_r}$, $\nu_r = S^{(1)}_{d_r,s_r}$, $\mathbf{h}_k = (H^{(1)}_{k,:})^\top$, and $\mathfrak{a}_r = \langle \mathbf{g}^{\mathrm{out}}, \mathbf{h}_{d_r} - \mathbf{e}_{[\mathbf{x}_0]_{s_r}}\rangle$. Row-centering does not change a directional derivative because $\mathbf{G}_1\mathbf{1}_T = 0$. The graph-row gradients have no query-backpropagation term, giving

$$-\langle \mathbf{G}_1, P_{ds}\rangle_F = \sum_{r=1}^{m} \sigma_r\nu_r\mathfrak{a}_r. \tag{129}$$

Write $\sigma_\star = \sigma_{r_\star}$ and $\nu_\star = \nu_{r_\star}$. The predecessor token y_1 occurs at $s^\star$ and the preceding destination. In the terminal row $d^\star$, that preceding destination and any non-target arm's terminal destination are distinct unpaired keys with the same content and positional scores. This remains true when a_2 is revealed. Hence

$$H^{(1)}_{d^\star,y_1} \le \nu_\star + (1-\nu_\star)/2, \qquad \mathfrak{a}_{r_\star} \ge \frac{C_1}{2}(1-\nu_\star). \tag{130}$$

Since $p_{y_1} \le e^\alpha/(e^\alpha + K - 1)$, every wrong-node or mask coordinate of $\mathbf{g}^{\mathrm{out}}$ exceeds its correct-node coordinate by at least $\alpha(1-p_{y_1}) \ge C_1$. Averaging these coordinate differences against the row $H^{(1)}_{d^\star,:}$ proves the second inequality. For any other row, $|\nu_r\mathfrak{a}_r| \le 2\alpha$. Meanwhile, Proposition 3.4, with equality contrast $\alpha_2 - \beta_2 = b$, gives, whenever $H^{(1)}_{a_1,v^\star} \ge b_0$,

$$\sum_{r\ne r_\star} \sigma_r \le \sigma_\star\mu_{\mathrm{match}}, \qquad \mu_{\mathrm{match}} := (m-1)e^{-b(2b_0-1)}. \tag{131}$$

Substitution into the directional derivative yields

$$-\langle \mathbf{G}_1, P_{ds}\rangle_F \ge \sigma_\star\left[\frac{C_1}{2}\nu_\star(1-\nu_\star) - 2\alpha\mu_{\mathrm{match}}\right]. \tag{132}$$

This bound controls all nonterminal graph-row contributions without a pool-mass lower-bound assumption.

By continuity, choose $\delta_{\mathrm{loc}} > 0$ such that on the closed radius-δ_{loc} ball about ϑ_0 in the trainable affine space,

$$H^{(1)}_{a_1,v^\star} \ge b_0, \qquad \sigma_\star \ge \sigma_{\min}/2, \qquad \nu_\star \ge \nu_{\min}/2, \qquad 1-\nu_\star \ge \omega_{\min}/2 \tag{133}$$

hold separately in both predecessor states. Finite population size and symmetry make these bounds simultaneous over inputs. Let

$$B_{\mathrm{flow}} := \max_{\|\vartheta-\vartheta_0\|_F \le \delta_{\mathrm{loc}}} \left(\sum_{\ell=1}^{2} \|\operatorname{Proj}_{\mathcal{U}_\ell} \nabla_{W_p^{(\ell)}} \mathcal{J}_P(\vartheta)\|_F^2\right)^{1/2}, \qquad \tau := \frac{\delta_{\mathrm{loc}}}{2B_{\mathrm{flow}}}. \tag{134}$$

The maximum is finite by smoothness and compactness, and positive by the descent estimate below at ϑ_0. A first-exit argument gives $\|\vartheta(t) - \vartheta_0\|_F \le B_{\text{flow}}(t - t_0) \le \delta_{\text{loc}}/2$ for $t \in [t_0, t_0 + \tau]$, so the solution remains in this neighborhood.

On that interval, the content-scale choice and net descent bound give

$$-\langle \mathbf{G}_1, P_{ds} \rangle_F \ge \frac{\sigma_{\min}}{2} \left[\frac{C_1 \nu_{\min} \omega_{\min}}{8} - \frac{C_1 \nu_{\min} \omega_{\min}}{16} \right] = \lambda_P. \tag{135}$$

Averaging over both predecessor tasks and the population preserves this lower bound. Since $P_{ds} \in \mathcal{U}_1$, Cauchy–Schwarz implies

$$\| \operatorname{Proj}_{\mathcal{U}_1} \nabla_{W_p^{(1)}} \mathcal{J}_P \|_F^2 \ge \frac{|\langle \nabla_{W_p^{(1)}} \mathcal{J}_P, P_{ds} \rangle_F|^2}{N_{ds}} \ge \frac{\lambda_P^2}{N_{ds}}. \tag{136}$$

The energy identity then gives $\dot{R}_P \le -c_P$. Thus

$$R_P(t) \le R_P(t_0) - c_P(t - t_0). \tag{137}$$

The admissible drop is strictly below $c_P \tau$. Continuity gives the hitting-time bound, with $R_P(t_P) = \eta$. □

Choice of constants and the staged threshold. At a fixed Layer-1 state, the query mass $H^{(1)}_{a_1, v^\star}$ in each predecessor state, $\nu_{\min}$, and $\omega_{\min}$ do not depend on the Layer-2 scale b. Thus the finite-scale condition is feasible, for example

$$b \ge \frac{1}{2b_0 - 1} \log \frac{32\alpha(m-1)}{C_1 \nu_{\min} \omega_{\min}}. \tag{138}$$

After making that choice, evaluate $\sigma_{\min}, \delta_{\text{loc}}, B_{\text{flow}}, c_P, \tau$ and the endpoint-entry conditions at the resulting state. These quantities can depend on the choice of b and on t_0; the proof does not assume their independence.

The circuit certificate uses an admissible staged threshold with $R_E(t) \le \eta$ on $[t_0, t_P]$. This threshold need not equal the early joint-training threshold: Stage I supplies the qualifying state, and Stage II proves a positive local risk drop from it. Neither arbitrary near-zero risk nor entry readiness follows by taking the early drop $h \to 0$.

B.6 Proofs for the reverse-circuit certificate

All statements here concern the both-masked input. The obstruction uses only the standing symmetries; endpoint provenance additionally uses $\alpha_1 - \beta_1 \le 0$. Stage II supplies $\alpha_1 - \beta_1 = -a < 0$ and $\alpha_2 - \beta_2 = b > 0$.

Lemma B.5 (Quantitative obstruction to non-reverse predecessor retrieval)**.** *Every population input satisfies*

$$1 - S^{(2)}_{a_1, d_{r_\star}} S^{(1)}_{d_{r_\star}, s_{r_\star}} \le \frac{\varepsilon_1}{\delta_L}. \tag{139}$$

Proof. Set $n = L - 1$ and retain the arm labels $h \in [R]$, the target arm $h_\star$, and the node labels $v_{h,j}$ of Section 2. Let s^h_j, d^h_j denote the serialized positions of edge $(v_{h,j-1}, v_{h,j})$. Thus $s^{h_\star}_n = s_{r_\star} = s^\star$ and $d^{h_\star}_n = d_{r_\star} = d^\star$. The predecessor token $y_1 = v_{n-1}$ occurs only at $s^\star$ and $d^{h_\star}_{n-1}$.

For either masked query a_i, there are constants $c_{i,S}, c_{i,D}$ such that $S^{(1)}_{a_i, s_r} = c_{i,S}$ and $S^{(1)}_{a_i, d_r} = c_{i,D}$ for all r. Every internal node occurs once in each graph role, and therefore $H^{(1)}_{a_i, v} = c_{i,S} + c_{i,D}$ for every internal node v. For a node key v, the Layer-2 content logit is a key-independent constant plus $(\alpha_2 - \beta_2) H^{(1)}_{a_i, v}$. Moreover, $\mathbf{W}^{(2)}_p \in \mathcal{V}_2$ makes the effective positional columns identical within each graph role. Hence Layer-2 attention is constant within each of the classes

$$\mathcal{C}^0_S = \{ s^h_1 : h \in [R] \}, \tag{140}$$

$$\mathcal{C}^I_S = \{ s^h_j : h \in [R],\ 2 \le j \le n \}, \tag{141}$$

$$\mathcal{C}^I_D = \{ d^h_j : h \in [R],\ 1 \le j < n \}, \tag{142}$$

together with the terminal singletons $\{d_n^h\}$ and the four special positions.

For every non-target arm $h \neq h_\star$, the table gives witness rows and distinct wrong keys j_1, j_2 matching the scores of $s_n^{h_\star}, d_{n-1}^{h_\star}$, respectively:

$\mathcal{C}$	k	j_1	j_2
$\mathcal{C}_S^0$	s_1^h	s_n^h	$d_n^{h_\star}$
$\mathcal{C}_S^I$	s_n^h	$s_1^{h_\star}$	$d_n^{h_\star}$
$\mathcal{C}_D^I$	d_{n-1}^h	$s_1^{h_\star}$	$d_n^{h_\star}$
$\{d_n^h\}$	d_n^h	$s_1^{h_\star}$	$d_n^{h_\star}$
$\{k\},\ k \in \mathcal{I}_{\mathrm{sp}}$	k	s_n^h	d_{n-1}^h

(143)

In the first four rows, each correct key and competitor have the same unpaired role and both differ from the query token. In the last row they have the same role and equality relation with the special query. The wrong keys differ from both correct occurrences. Hence each witness satisfies $H_{k,y_1}^{(1)} \leq 1/2$. Each of $\mathcal{C}_S^0, \mathcal{C}_S^I, \mathcal{C}_D^I$ contains $R-1$ distinct witnesses and has size at most $R(n-1)$. Therefore

$$\frac{1}{|\mathcal{C}|}\sum_{k\in\mathcal{C}} H_{k,y_1}^{(1)} \leq 1 - \frac{R-1}{2|\mathcal{C}|} \leq 1 - \frac{R-1}{2R(n-1)} \leq 1 - \delta_L. \tag{144}$$

Every remaining singleton except $\{d_n^{h_\star}\}$ has value at most $1/2 \leq 1 - \delta_L$. All classes except $\{d_n^{h_\star}\}$ therefore share this deficit.

For query $d^\star$, the keys $d_{n-1}^{h_\star}$ and d_n^h of a non-target arm h are unpaired destinations different from the query token and hence have identical scores. It follows that

$$H_{d^\star,y_1}^{(1)} \leq S_{d^\star,s^\star}^{(1)} + \frac{1 - S_{d^\star,s^\star}^{(1)}}{2}. \tag{145}$$

Using the constant Layer-2 weights within each remaining class gives

$$H_{a_1,y_1}^{(2)} \leq S_{a_1,d^\star}^{(2)} \left(S_{d^\star,s^\star}^{(1)} + \frac{1 - S_{d^\star,s^\star}^{(1)}}{2} \right) + (1 - S_{a_1,d^\star}^{(2)})(1 - \delta_L), \tag{146}$$

$$\varepsilon_1 \geq \frac{1}{2} S_{a_1,d^\star}^{(2)} (1 - S_{d^\star,s^\star}^{(1)}) + \delta_L (1 - S_{a_1,d^\star}^{(2)}) \tag{147}$$

$$\geq \delta_L \left[S_{a_1,d^\star}^{(2)} (1 - S_{d^\star,s^\star}^{(1)}) + 1 - S_{a_1,d^\star}^{(2)} \right] \tag{148}$$

$$= \delta_L \left(1 - S_{a_1,d^\star}^{(2)} S_{d^\star,s^\star}^{(1)} \right), \tag{149}$$

which is the claim. □

Lemma B.6 (Joint supervision constrains endpoint provenance)**.** *Every population input satisfies*

$$1 - Q_{a_2,q_t} \leq 3\varepsilon_2 + \frac{\varepsilon_1}{\delta_L}. \tag{150}$$

Proof. The endpoint occurrence bound and $1 - \varepsilon_2 = Q_{a_2,q_t} + Q_{a_2,d^\star}$ give

$$1 - Q_{a_2,q_t} \leq 3\varepsilon_2 + S_{a_2,d^\star}^{(2)} S_{d^\star,d^\star}^{(1)}. \tag{151}$$

Finally, Lemma B.5 implies

$$S_{a_2,d^\star}^{(2)} S_{d^\star,d^\star}^{(1)} \leq 1 - S_{d^\star,s^\star}^{(1)} \leq \frac{\varepsilon_1}{\delta_L}. \tag{152}$$

Combining the two inequalities proves the claim. Unlike the selected-a bound in Lemma B.2, this version uses the attained predecessor purity and needs no numerical bound on B_{self}. □

Proof of Theorem 3.5. Stage II preserves the standing fixed spaces: its content is symmetry-fixed and row-centered, and its trainable positional subspaces lie in $\mathcal{V}_p$ and $\mathcal{V}_2$. Since $\alpha_1 - \beta_1 = -a < 0$, both structural lemmas apply throughout $[t_0, t_P]$.

Lemma B.4 supplies $t_P < \infty$ and $R_P(t_P) \le \eta$. The prescribed endpoint stability gives $R_E(t_P) \le \eta$, so

$$\mathcal{R}(t_P) = \frac{1}{4}\big(R_E(t_P) + R_P(t_P)\big) \le \frac{\eta}{2} = \rho. \tag{153}$$

Consequently $t_\rho \le t_P < \infty$, and continuity gives $\mathcal{R}(t_\rho) \le \rho$.

Fix $t \in [t_0, t_P]$ and a both-masked population input. The loss–purity bound gives $C_1(\varepsilon_1 + \varepsilon_2) \le 4\mathcal{R}(t)$ uniformly over the population. Lemma B.5 gives

$$1 - S^{(2)}_{a_1,d^\star} S^{(1)}_{d^\star,s^\star} \le \frac{\varepsilon_1}{\delta_L} \le \frac{4}{C_1\delta_L}\mathcal{R}(t) = C_2\mathcal{R}(t). \tag{154}$$

Since $\delta_L^{-1} = 4(L-2) \ge 4 > 3$, Lemma B.6 gives

$$1 - Q_{a_2,q_t} \le 3\varepsilon_2 + \frac{\varepsilon_1}{\delta_L} \le \frac{\varepsilon_1 + \varepsilon_2}{\delta_L} \le C_2\mathcal{R}(t). \tag{155}$$

Bounding the two impurities jointly, rather than each by $4\mathcal{R}/C_1$, yields the common constant C_2.

Evaluate both bounds at t_ρ and t_P, where $\mathcal{R} \le \rho$. Each predecessor factor is no smaller than their product. Lemma 3.3 gives endpoint-route mass through $\mathcal{A}$ at least $1 - C_2\rho/\varepsilon$; carrier purity holds by definition, and $\mathcal{A} \neq \varnothing$ when $C_2\rho < \varepsilon$. If $C_2\rho \le \varepsilon^2$, the respective route masses are at least $1 - \varepsilon^2$ and $1 - \varepsilon$. □

C Proofs for Section 4

We retain the stage and symmetry setting of Section 4, the constant C_1, and the loss–purity bound of Lemma B.1. As in Appendix B, write $\mathcal{L}(e;t), R(e;t)$ for conditional loss and excess risk, and $\mathsf{E}(e;t)$ for the conditional success event; omit t when fixed. The targets are $y_1 = v_{L-2}$ and $y_2 = v^\star = v_{L-1}$. We specify each input before evaluating forward quantities, without mask-state subscripts.

C.1 Decoding Strategies

We denote the partially decoded sequence after $h \in \{0,1,2\}$ updates by $\mathbf{x}^{(h)}$ with $\mathbf{x}^{(0)} = \mathbf{x}_{\inf}$. Define the set of currently masked answer indices, token probabilities, and greedy predictions by

$$\mathcal{M}^{(h)} = \{j \in \{1,2\} : [\mathbf{x}^{(h)}]_{a_j} = [\texttt{MASK}]\} \tag{156}$$

$$p_j(v \mid \mathbf{x}^{(h)}) = [\mathbf{\Pi}(\mathbf{x}^{(h)})]_{a_j,v}, \tag{157}$$

$$\hat{y}_j(\mathbf{x}^{(h)}) \in \arg\max_{v\in\mathcal{V}} p_j(v \mid \mathbf{x}^{(h)}). \tag{158}$$

The fixed tie-breaking rule selects a single maximizer. After selecting an index $j \in \mathcal{M}^{(h)}$, the decoding state is updated as $\mathbf{x}^{(h+1)} = (\mathbf{x}^{(h)})^{a_j \leftarrow \hat{y}_j}$. For each j, let $p_j^{(1)}$ and $p_j^{(2)}$ denote the largest and second-largest values, among $\{p_j(v \mid \mathbf{x}^{(h)}) : v \in \mathcal{V}\}$.

- **Fixed LTR/RTL.** Decode the answer positions in the fixed order (a_1, a_2) for LTR or (a_2, a_1) for RTL.
- **Random.** At each decoding step, sample $j \sim \mathrm{Unif}(\mathcal{M}^{(h)})$.
- **Top Probability.** At each decoding step, select $j^\star \in \arg\max_{j\in\mathcal{M}^{(h)}} \max_{v\in\mathcal{V}} p_j(v \mid \mathbf{x}^{(h)})$.
- **Top Margin.** At each decoding step, select $j^\star \in \arg\max_{j\in\mathcal{M}^{(h)}} \big[p_j^{(1)} - p_j^{(2)}\big]$.
- **Immediate Oracle.** At each decoding step, select $j^\star \in \arg\max_{j\in\mathcal{M}^{(h)}} p_j(y_j \mid \mathbf{x}^{(h)})$, where $y_j = [\mathbf{x}_0]_{a_j}$ is the ground-truth answer token at position a_j.
- **Full-Value Oracle.** For $o \in \{\mathrm{LTR}, \mathrm{RTL}\}$, let $\mathrm{Rollout}(\mathbf{x}^{(0)}; o)$ denote the final answer pair obtained by greedy decoding in order o. Select

$$o^\star \in \arg\max_{o\in\{\mathrm{LTR},\mathrm{RTL}\}} \mathbf{1}\Big\{\mathrm{Rollout}(\mathbf{x}^{(0)}; o) = (y_1, y_2)\Big\},$$

  breaking ties in favor of LTR.

Exact success identities. For deterministic greedy predictions with a fixed tie-breaking rule, a two-step rollout succeeds if and only if its first token is correct and its second token is correct in the resulting state. When the first token is correct, that state is exactly the ground-truth-conditioned state of the corresponding single-masked problem. This proves equation 20 without a teacher-forcing approximation to the success event.

C.2 Before Endpoint Readiness

Corollary C.1 (Initial decoding failure and equality)**.** *Consider the zero-initialized population joint flow of Theorem 3.1 on $G(R,L)$, with $R \geq 2$ and $L \geq 3$. There is $\varepsilon_{\mathrm{init}} > 0$ such that $\mathrm{Acc}_{\mathrm{RTL}}(t) = \mathrm{Acc}_{\mathrm{LTR}}(t) = 0$ for $0 \leq t < \varepsilon_{\mathrm{init}}$. For every sufficiently small $h > 0$ in that theorem, put $\eta = R_E(0) - h$ and $t_E = \tau_E(\eta)$. Then $t_E < \min\{\varepsilon_{\mathrm{init}}, \tau_P(\eta)\}$, so the same zero-accuracy equality holds on the nonempty interval $t_E < t < \min\{\varepsilon_{\mathrm{init}}, \tau_P(\eta)\}$.*

Proof of Corollary C.1. At zero initialization on the both-masked input, both attention matrices equal $T^{-1}\mathbf{1}_T\mathbf{1}_T^\top$, so

$$\mathbf{H}^{(2)}(0) = \frac{1}{T}\mathbf{1}_T\mathbf{1}_T^\top\mathbf{Z}. \tag{159}$$

The root u occurs $R+1$ times (at the first source of each arm and at q_s), whereas each answer occurs twice. With $y_i = [\mathbf{x}_0]_{a_i}$,

$$H^{(2)}_{a_i,u}(0) - H^{(2)}_{a_i,y_i}(0) = \frac{R-1}{T} > 0, \qquad i \in \{1,2\}. \tag{160}$$

The fixed head preserves this ordering. Continuity and the finite population give a common $\varepsilon_{\text{init}} > 0$ on which both possible first-token predictions are wrong. The success identities equation 20 give

$$\text{Acc}_{\text{RTL}}(t) = \text{Acc}_{\text{LTR}}(t) = 0, \qquad 0 \le t < \varepsilon_{\text{init}}, \tag{161}$$

regardless of ties among wrong nodes.

The proof of Theorem 3.1 gives $c_E = -\dot{R}_E(0) > 0$ and $t_E \le 2h/c_E$ for all sufficiently small $h > 0$; see equation 102. Decreasing h if necessary yields $t_E < \varepsilon_{\text{init}}$, while $t_E < \tau_P(\eta)$ is the theorem's conclusion. This proves the remaining assertion. □

Remark C.2. Early endpoint-risk separation does not by itself imply an RTL accuracy advantage: both decoding orders can still have zero rollout accuracy. Thus, connecting endpoint-first learning to reverse-order decoding requires an additional bootstrap or circuit-formation argument.

C.3 Stage II: Endpoint-Reveal Bootstrap

Inputs, margins, and the distinguished path. All comparisons in this subsection use finite states with the selected content blocks of Stage II, $a, b > 0$, and its symmetry and Layer-2 pair-blindness conditions. Fix one clean sequence $\mathbf{x}_0$ and retain $s^\star = s_{r_\star}$ and $d^\star = d_{r_\star}$, with

$$y_1 = [\mathbf{x}_0]_{a_1} = [\mathbf{x}_0]_{s^\star} = v_{L-2}, \qquad v^\star = [\mathbf{x}_0]_{a_2} = [\mathbf{x}_0]_{d^\star} = v_{L-1}. \tag{162}$$

Unprimed quantities are evaluated with both answers masked. Primed quantities are evaluated after revealing the correct endpoint $a_2 = v^\star$ while leaving a_1 masked, at exactly the same parameter state. Thus

$$\widetilde{\mathbf{x}} = \mathbf{x}_0[\{a_1, a_2\}], \qquad \widetilde{\mathbf{x}}' = \mathbf{x}_0[\{a_1\}]. \tag{163}$$

Let $\mathbf{Z}, \mathbf{Z}'$ be their one-hot token matrices, including the mask coordinate. For the unprimed input,

$$\mathbf{H}^{(1)} = \mathbf{S}^{(1)}\mathbf{Z}, \qquad \mathbf{H}^{(2)} = \mathbf{S}^{(2)}\mathbf{S}^{(1)}\mathbf{Z}; \tag{164}$$

the same identities with primes define $\mathbf{H}^{(1)\prime}$ and $\mathbf{H}^{(2)\prime}$. Define the route mass and destination-pool exposure by

$$r = S^{(2)}_{a_1,d^\star} S^{(1)}_{d^\star,s^\star}, \qquad \chi = \sum_{j=1}^{m} S^{(2)}_{a_1,d_j}, \tag{165}$$

and let r', χ' denote the same quantities for the primed input. The route mass is one scalar path weight, not the full matrix product $\mathbf{S}^{(2)}\mathbf{S}^{(1)}$. Since $s^\star$ always contains y_1, its contribution to every correct-versus-wrong output difference is exactly $+r$, or $+r'$ after reveal. The corresponding predecessor margins are

$$M(t) = H^{(2)}_{a_1,y_1}(t) - \max_{v \in \mathcal{V}\setminus\{y_1\}} H^{(2)}_{a_1,v}(t), \tag{166}$$

$$M'(t) = H^{(2)\prime}_{a_1,y_1}(t) - \max_{v \in \mathcal{V}\setminus\{y_1\}} H^{(2)\prime}_{a_1,v}(t). \tag{167}$$

These are the two margins of Section 4.1, with the same unprimed/primed input convention. They minimize over node tokens only, excluding the mask. Multiplication by the fixed head scale $\alpha > 0$ preserves their signs.

Interference is a signed output-mass contribution, not risk. The only additional scalar in the theorem is ε_{int}. For $[z]_+ = \max\{z, 0\}$, define it on this clean sequence by the full non-route path sum

$$\varepsilon_{\text{int}}(t) := \max_{v \in \mathcal{V}\setminus\{y_1\}} \Bigg[\sum_{k,j \in [T](k,j) \neq (d^\star, s^\star)} \Bigg\{ S^{(2)}_{a_1,k}(t) S^{(1)}_{k,j}(t) \left(\mathbf{1}_{\{[\widetilde{\mathbf{x}}]_j = y_1\}} - \mathbf{1}_{\{[\widetilde{\mathbf{x}}]_j = v\}} \right) \tag{168}$$

$$- S^{(2)\prime}_{a_1,k}(t) S^{(1)\prime}_{k,j}(t) \left(\mathbf{1}_{\{[\widetilde{\mathbf{x}}']_j = y_1\}} - \mathbf{1}_{\{[\widetilde{\mathbf{x}}']_j = v\}} \right) \Bigg\} \Bigg]_+ . \tag{169}$$

For each wrong node, the sum is its non-route correct-versus-wrong contribution before reveal minus that after reveal. The positive part retains only adverse changes; the maximum covers every possible competitor. Other paths that end at another occurrence of y_1 are included with a positive sign, and paths carrying the wrong node with a negative sign. The changed value at a_2 is included through $\widetilde{\mathbf{x}}'$, rather than treated as a fixed input. Thus ε_{int} and r, r' are dimensionless output-mass quantities, not conditional excess risks.

This definition does not assume that ε_{int} is small. It is an evaluated forward-pass quantity, not an independent parameter or a gradient remainder. A verified upper bound may replace it in the bootstrap dominance condition. Under exact symmetry, relabeling node identities and permuting serialized edges permute the summands and the maximization set, preserving the distinguished terminal path. Hence r, r', M, M', and ε_{int} are constant on the clean-sequence orbit, just as the elementary loss is in Lemma B.1. Checking a representative sequence suffices only under this exact symmetry.

Proof of Lemma 4.1. In the Layer-1 row whose query is the masked a_1, changing only the a_2 key from mask to $v^\star$ increases that key's content score by exactly a; all other scores in this row are unchanged. To display the normalization explicitly, use the temporary abbreviation $w := S^{(1)}_{a_1,a_2} > 0$. The new denominator is the old denominator times $1 + (e^a - 1)w$. Consequently

$$H^{(1)\prime}_{a_1,v^\star} = \frac{H^{(1)}_{a_1,v^\star} + e^a w}{1 + (e^a - 1)w}, \tag{170}$$

$$H^{(1)\prime}_{a_1,v} = \frac{H^{(1)}_{a_1,v}}{1 + (e^a - 1)w}, \qquad v \in \mathcal{V} \setminus \{v^\star\}. \tag{171}$$

Subtraction gives

$$H^{(1)\prime}_{a_1,v^\star} - H^{(1)}_{a_1,v^\star} = \frac{w\left[e^a - (e^a - 1)H^{(1)}_{a_1,v^\star}\right]}{1 + (e^a - 1)w} > 0, \tag{172}$$

because $0 \le H^{(1)}_{a_1,v^\star} \le 1$. Every other node value in this query is nonincreasing. Since each non-target destination token is distinct from $v^\star$, for every $r \neq r_\star$,

$$H^{(1)\prime}_{a_1,v^\star} - H^{(1)\prime}_{a_1,[\mathbf{x}_0]_{d_r}} > H^{(1)}_{a_1,v^\star} - H^{(1)}_{a_1,[\mathbf{x}_0]_{d_r}}. \tag{173}$$

The positional score cancels between destination columns by Layer-2 pair blindness. In the notation of Lemma 2.1, the equality contrast is $\alpha_2 - \beta_2 = b > 0$. For the unprimed input, the exact identity is

$$\frac{S^{(2)}_{a_1,d^\star}}{\chi} = \left[1 + \sum_{r \neq r_\star} e^{-b(H^{(1)}_{a_1,v^\star} - H^{(1)}_{a_1,[\mathbf{x}_0]_{d_r}})}\right]^{-1}. \tag{174}$$

the same identity with primes holds after reveal. Substituting the strict gap comparison proves the strict matching comparison. No assumption that the query endpoint mass exceeds $1/2$ is needed for this comparison.

Finally the Layer-1 query at $d^\star$ is $v^\star$, for which the selected content gives the same score to a mask key and a $v^\star$ key. Revealing a_2 leaves all scores in this row unchanged, proving the attention-row equality. In particular,

$$r' - r = S^{(1)}_{d^\star,s^\star}\left(S^{(2)\prime}_{a_1,d^\star} - S^{(2)}_{a_1,d^\star}\right). \tag{175}$$

This is a change in utilization of the same paired-source retrieval step, not a comparison between two separately trained models. □

What the matching lemma does not prove. The denominators χ' and χ need not agree. Therefore, the improved within-pool matching alone does not imply either $r' > r$ or a margin ordering. Revealing a_2 also creates a same-content Layer-2 key competing with $d^\star$. Its effect, and every other non-route change, is retained in equation 169. The bootstrap dominance condition controls these changes beyond within-pool matching.

Lemma C.3 (Route dominance implies a quantitative margin advantage)**.** *At any finite Stage II state, compare the both-masked and endpoint-revealed inputs at the same parameters. With r, r', M, M' defined above and ε_{int} as in equation 169,*

$$M' - M \geq r' - r - \varepsilon_{\text{int}}. \tag{176}$$

In particular, $r' - r > \varepsilon_{\text{int}}$ implies $M' > M$, without fixing the largest wrong node.

Proof. On the both-masked input, expanding both attention layers gives, for every node v,

$$H^{(2)}_{a_1,v} = \sum_{k=1}^{T}\sum_{j=1}^{T} S^{(2)}_{a_1,k} S^{(1)}_{k,j} \mathbf{1}_{\{[\widetilde{\mathbf{x}}]_j = v\}}. \tag{177}$$

Since $[\widetilde{\mathbf{x}}]_{s^\star} = [\widetilde{\mathbf{x}}']_{s^\star} = y_1$, for every $v \neq y_1$ we have the exact path decomposition

$$H^{(2)}_{a_1,y_1} - H^{(2)}_{a_1,v} = r + \sum_{\substack{k,j\in[T]\\(k,j)\neq(d^\star,s^\star)}} S^{(2)}_{a_1,k} S^{(1)}_{k,j} \left(\mathbf{1}_{\{[\widetilde{\mathbf{x}}]_j = y_1\}} - \mathbf{1}_{\{[\widetilde{\mathbf{x}}]_j = v\}}\right). \tag{178}$$

The same decomposition with primes holds after reveal. Subtracting the unprimed identity from the primed one and using equation 169 yields, for each wrong node,

$$\begin{aligned}\left(H^{(2)\prime}_{a_1,y_1} - H^{(2)\prime}_{a_1,v}\right) - \left(H^{(2)}_{a_1,y_1} - H^{(2)}_{a_1,v}\right)\\ \geq r' - r - \varepsilon_{\text{int}}.\end{aligned} \tag{179}$$

Hence every revealed-input correct-versus-wrong difference is at least

$$M + r' - r - \varepsilon_{\text{int}}. \tag{180}$$

Taking the minimum over $v \in \mathcal{V}\backslash\{y_1\}$ on the primed side proves equation 176. All path contributions and changing competitors are retained. This is a forward-pass bound, not a claim about gradient dynamics. □

Proof of Theorem 4.2. In the input convention above, the theorem assumes $M(t_0) < 0$ and finite first contact $t_\star = \inf\{t \geq t_0 : M(t) \geq 0\}$, together with $H^{(2)}_{a_2,v^\star}(t_\star) > 1/2$ and $r'(t_\star) - r(t_\star) > \varepsilon_{\text{int}}(t_\star)$. Continuity and $M(t_0) < 0$ imply $t_\star > t_0$. By the definition of $t_\star$,

$$M(t) < 0, \qquad t_0 \leq t < t_\star, \qquad\qquad M(t_\star) = 0. \tag{181}$$

To justify the equality, a strictly negative value at $t_\star$ would contradict that it is the infimum of the nonnegative set, while a strictly positive value would give an earlier nonnegative time by continuity. Lemma C.3 and the strict dominance assumption give

$$M'(t_\star) \geq r'(t_\star) - r(t_\star) - \varepsilon_{\text{int}}(t_\star) > 0. \tag{182}$$

Both this margin and the unprimed endpoint value are continuous. Choose $0 < \delta < t_\star - t_0$ small enough that

$$M'(t) > 0, \qquad\qquad H^{(2)}_{a_2,v^\star}(t) > \frac{1}{2}, \qquad t \in (t_\star - \delta, t_\star). \tag{183}$$

Equation equation 181 makes $M(t) < 0$ on the same interval. Thus $M' > 0 > M$, proving strict margin advantage as well as opposite correctness signs.

The endpoint bound implies that $v^\star$ is the unique greedy prediction at a_2 on the both-masked input. RTL therefore produces the endpoint-revealed input, whose predecessor prediction is uniquely correct. LTR is wrong on its first token because $M < 0$. These statements hold on the entire population by exact symmetry. Substitution into equation 20 gives

$$\text{Acc}_{\text{RTL}}(t) = \mathbb{P}\big(\mathsf{E}(2 \mid \varnothing; t) \cap \mathsf{E}(1 \mid 2; t)\big) = 1, \tag{184}$$

$$\text{Acc}_{\text{LTR}}(t) = \mathbb{P}\big(\mathsf{E}(1 \mid \varnothing; t) \cap \mathsf{E}(2 \mid 1; t)\big) = 0. \tag{185}$$

No correctness assumption on $2 \mid 1$ is required. The open interval excludes the boundary tie at $t_\star$, and all claimed predictions are independent of tie-breaking. □

Relation to stage-wise learning. The matching lemma follows from the Stage II setting. Route dominance additionally controls all wrong nodes and does not follow from endpoint readiness. The endpoint correctness and finite boundary contact are also bootstrap premises. An attained endpoint risk $R_E < C_1/2$ suffices for the endpoint value bound, and an actually attained predecessor risk $R_P < C_1/2$ after t_0 suffices for finite $t_\star$, by Lemma B.1. Local Stage II attainment does not automatically reach these levels.

The first contact is chosen for the unconditional predecessor, guaranteeing that LTR was still wrong immediately beforehand. The proof needs neither margin monotonicity nor a no-return property. It establishes an interval of bootstrap advantage under the stated conditions, not their attainment by the learning dynamics.

C.4 From Circuit Separation to Reverse Confidence

Proof of Proposition 4.3. Fix a symmetry-fixed, Layer-2 pair-blind state satisfying the circuit-separation premise of Proposition 4.3. Use the both-masked input throughout this proof and write $h_{i,v} := H^{(2)}_{a_i,v}$ and $y_i = [\mathbf{x}_0]_{a_i}$. Since q_t carries y_2, nonnegativity gives $h_{2,y_2} \geq Q_{a_2,q_t} = 1 - \delta_Q$. The distinguished predecessor route carries y_1, while Lemma B.5 bounds its impurity. Thus

$$r \leq h_{1,y_1} \leq 1 - \delta_L(1-r), \qquad h_{1,v} \leq 1 - r \quad (v \neq y_1). \tag{186}$$

Define the resulting upper bound on any node value at a_1 and the endpoint's lower correct-versus-wrong gap by

$$B := 1 - \min\{r, \delta_L(1-r)\}, \qquad d := 1 - 2\delta_Q. \tag{187}$$

Then $\max_v h_{1,v} \leq B$ and $h_{2,y_2} - h_{2,v} \geq d$ for every wrong node v. The circuit-separation premise is exactly $d > B \geq 0$.

Define the strictly increasing functions on $z \geq 0$,

$$F_{\mathrm{prob}}(z) := \frac{1}{1 + (K-1)e^{-\alpha z}}, \qquad F_{\mathrm{margin}}(z) := \frac{1 - e^{-\alpha z}}{1 + (K-1)e^{-\alpha z}}. \tag{188}$$

At a_2, all wrong-to-correct probability ratios are at most $e^{-\alpha d}$, so $c_2^{\mathrm{prob}} \geq F_{\mathrm{prob}}(d)$. At a_1, nonnegativity and $\max_v h_{1,v} \leq B$ make every gap from the largest node value at most B. Normalizing the probability ratios therefore gives $c_1^{\mathrm{prob}} \leq F_{\mathrm{prob}}(B)$.

The same bounds give $p_2^{(2)} \leq e^{-\alpha d} p_2^{(1)}$ and $p_1^{(2)} \geq e^{-\alpha B} p_1^{(1)}$. Consequently

$$c_2^{\mathrm{margin}} \geq (1 - e^{-\alpha d}) p_2^{(1)} \geq F_{\mathrm{margin}}(d), \tag{189}$$

$$c_1^{\mathrm{margin}} \leq (1 - e^{-\alpha B}) p_1^{(1)} \leq F_{\mathrm{margin}}(B). \tag{190}$$

Strict monotonicity and $d > B$ prove both confidence comparisons, regardless of the largest node at a_1. At any time in the bootstrap interval where this premise also holds, either confidence rule chooses a_2 first. Theorem 4.2 then makes both steps of that rollout correct, proving the combined accuracy conclusion in the main text. □

C.5 Final-Stage Order Independence and Confidence Agreement

Proof of Theorem 4.4. At a symmetry-fixed state, fix any conditional problem $e \in \mathcal{E}_4$ and evaluate the forward pass in that problem. Let $i \in \{a_1, a_2\}$ be its supervised position and $y = [\mathbf{x}_0]_i$ its target. Lemma B.1 gives

$$1 - H^{(2)}_{i,y} \leq \frac{R(e;t)}{C_1}. \tag{191}$$

Choose sufficiently small $0 < \eta < C_1/2$ so that the continuation hypothesis applies. For every $t \geq t_P(\eta)$, each nonnegative conditional excess is at most its group sum and hence at most η. All four conditional correct values exceed $1/2$, uniformly over the population. Either first-token choice is uniquely correct, and the remaining conditional prediction is correct as well. Every monotone greedy order therefore has accuracy one for all $t \geq t_P(\eta)$, proving the finite-time assertion and, in particular, its accuracy limit. No risk monotonicity beyond the stated continuation is needed.

For the confidence limits, return to the both-masked input and let $\mathbf{h}_i[\mathcal{V}] = (H^{(2)}_{a_i,v})_{v\in\mathcal{V}}$ be its node-value vector at position a_i, and put $\varepsilon_i = 1 - H^{(2)}_{a_i,y_i}$. At the pure correct value $\mathbf{e}_{y_i}$, the two confidence scores are independent of i and equal

$$p_* = \frac{e^\alpha}{e^\alpha + K - 1}, \qquad m_* = \frac{e^\alpha - 1}{e^\alpha + K - 1}. \tag{192}$$

They need not equal one at finite α. Nonnegativity gives $\|\mathbf{h}_i[\mathcal{V}] - \mathbf{e}_{y_i}\|_\infty \le \varepsilon_i$. Each row of the softmax Jacobian has absolute row sum $2p_v(1-p_v) \le 1/2$, so

$$\|\operatorname{softmax}(\alpha\mathbf{h}_i[\mathcal{V}]) - \operatorname{softmax}(\alpha\mathbf{e}_{y_i})\|_\infty \le \frac{\alpha}{2}\varepsilon_i. \tag{193}$$

The largest and second-largest coordinates are each 1-Lipschitz in infinity norm. Hence

$$|c_i^{\mathrm{prob}} - p_*| \le \frac{\alpha}{2}\varepsilon_i, \qquad |c_i^{\mathrm{margin}} - m_*| \le \alpha\varepsilon_i. \tag{194}$$

Combining the two positions and using equation 77 yields

$$|D_{\mathrm{prob}}| \le \frac{\alpha}{2}(\varepsilon_1 + \varepsilon_2) \le \frac{2\alpha}{C_1}\mathcal{R}, \tag{195}$$

$$|D_{\mathrm{margin}}| \le \alpha(\varepsilon_1 + \varepsilon_2) \le \frac{4\alpha}{C_1}\mathcal{R}. \tag{196}$$

The continuation hypothesis gives $\mathcal{R} = (R_E + R_P)/4 \le \eta/2$ for all $t \ge t_P(\eta)$, for every sufficiently small $\eta > 0$. Thus both confidence differences tend to zero, completing the proof. □

D EXPERIMENTAL DETAILS

D.1 MODEL TRAINING SETUP

The attention parameters are initialized independently as $\mathbf{W}_c^{(\ell)}, \mathbf{W}_p^{(\ell)} \sim \mathcal{N}(0, \sigma_0^2)$ where $\sigma_0 = 10^{-2}$. The output head is initialized to the node-token basis $\mathbf{U} = [\alpha \mathbf{I}_K \quad \mathbf{0}]$ where $\alpha = 5$. No causal mask is applied. The model has no residual connection, MLP, normalization, dropout, or diffusion-time embedding.

Optimization. We fix the output head at its initialization, then all parameters $\theta = \left\{\mathbf{W}_c^{(1)}, \mathbf{W}_p^{(1)}, \mathbf{W}_c^{(2)}, \mathbf{W}_p^{(2)}\right\}$ are jointly optimized using stochastic full-batch AdamW. At each update, clean sequences and mask patterns are independently sampled according to the distributions above. Optimization used $\eta_{\text{lr}} = 10^{-2}$, $(\beta_1, \beta_2) = (0.9, 0.95)$, zero weight decay, and $S = 100\text{k}$ steps with no learning-rate scheduler, warmup, gradient clipping, curriculum, or staged training.

D.2 $G(R, L)$ RESULTS WITH MINIBATCH SETTING

Using minibatches $B = 512$ and a train/test split breaks the exact data symmetry assumed in our theoretical analysis. Consequently, the resulting training dynamics need not follow the dynamics proved in the main text; we include these experiments as complementary evidence that the same reverse-circuit formation mechanism can emerge under symmetry-broken training.

Attention maps. We visualize train and test attention maps across the four mask states to assess whether reverse-retrieval structure persists under minibatch training.

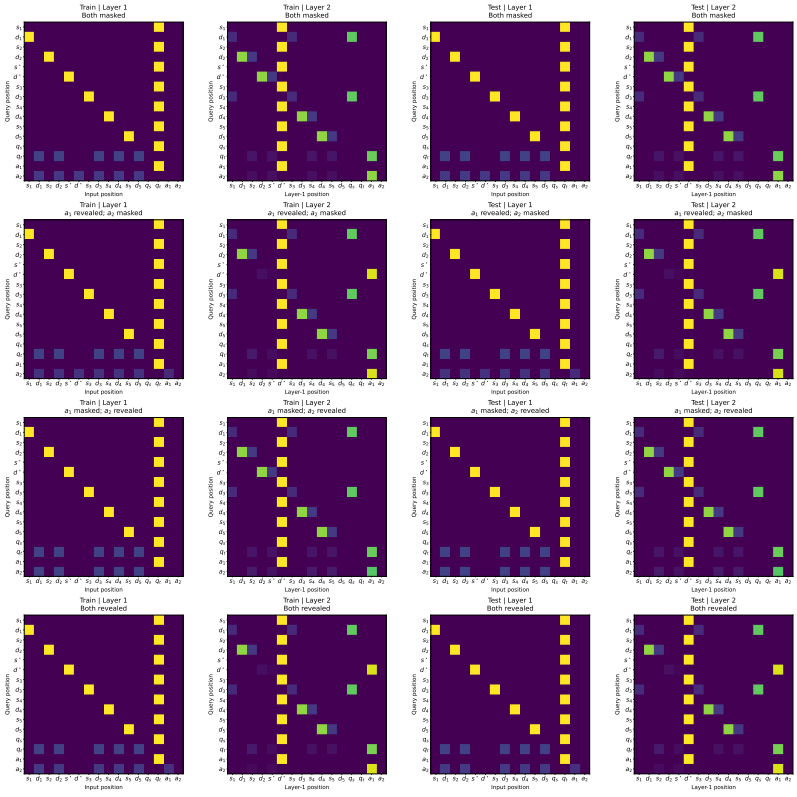


Figure 5: Attention maps for the minibatch experiment on $G(2, 4)$. The 4×4 grid reports train and test attention maps for the first and second layers across all four mask states.

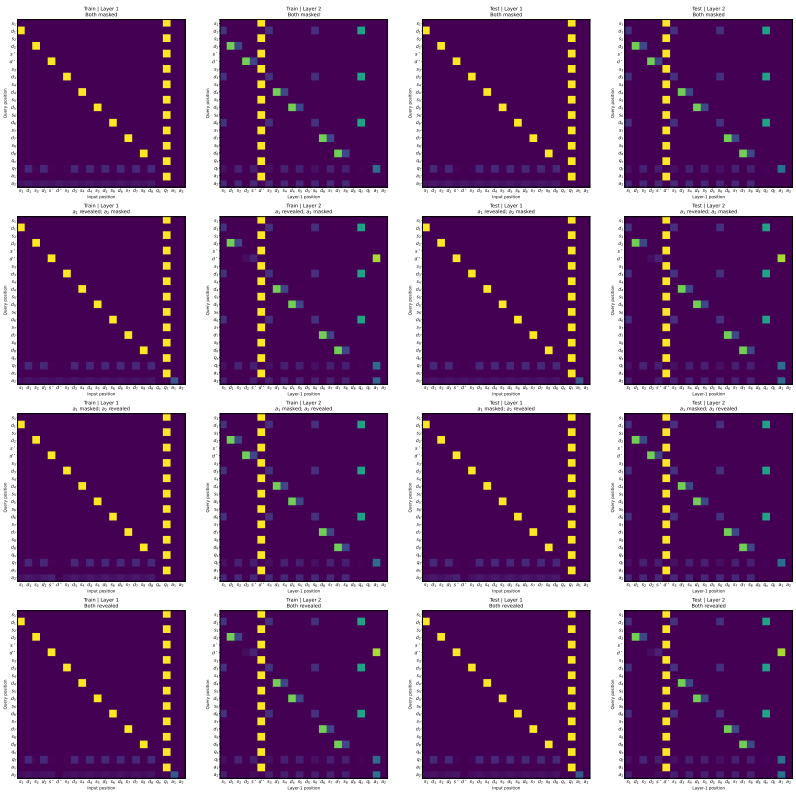


Figure 6: $G(3, 4)$ train and test attention map

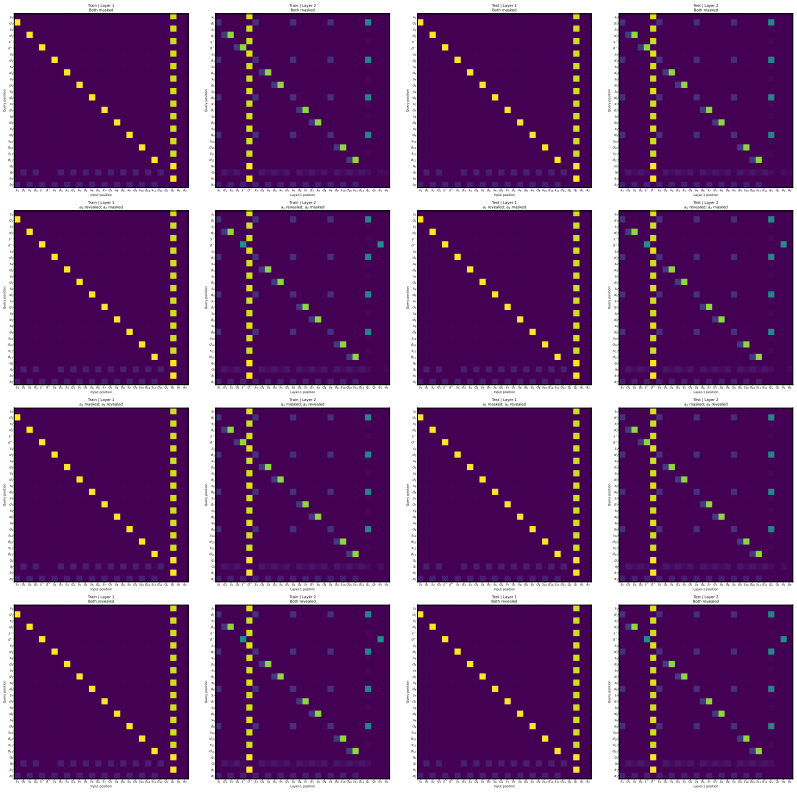


Figure 7: $G(3, 4)$ train and test attention map

Training dynamics. We additionally report the learning dynamics on the both-masked input. Each figure tracks train loss, test accuracy, train route mass, and the fraction of reverse choices among non-tied decoding decisions. Together, these quantities compare predictive performance with the emergence of the reverse-retrieval route during training.

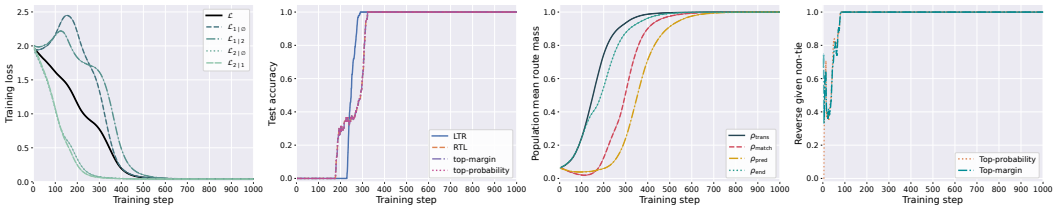


Figure 8: Training dynamics for $G(2, 4)$ on the both-masked input. The four panels report train loss, test accuracy, train route mass, and the fraction of reverse choices among non-tied decisions.

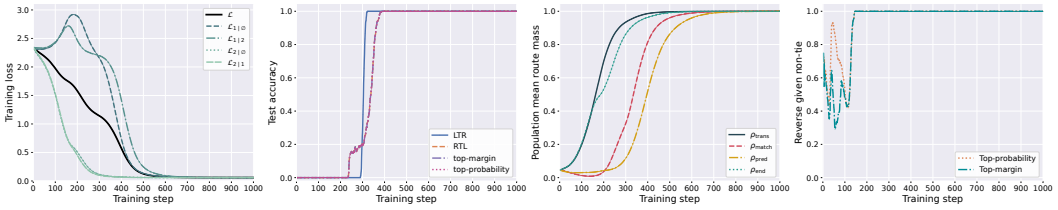


Figure 9: Training dynamics for $G(3, 4)$ on the both-masked input.

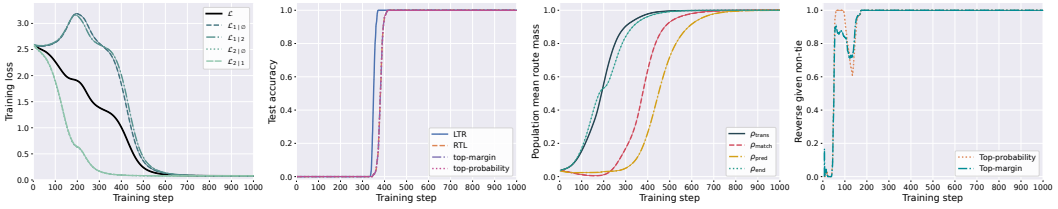


Figure 10: Training dynamics for $G(4, 4)$ on the both-masked input.