HIDDEN ROUTING: A TIGHT ALPHABET-SIZE GAP IN AUTOREGRESSIVE ONLINE LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

An autoregressive model generates a sequence by applying the same next-token rule at every step, and intermediate steps can carry information that is absent from the final token. Endpoint-only feedback observes only that final token, while full-trajectory feedback observes every step, so the two protocols can differ in online complexity even under the same prediction target and loss. For binary alphabets, existing analyses show how this gap grows with the rollout length, but they do not determine whether alphabet size adds a separate cost. We prove that this alphabet-size cost is unavoidable in the worst case over unrestricted next-token classes. For finite alphabets, a multiclass sequential-growth argument bounds final-token complexity linearly in the base-class Littlestone dimension and logarithmically in the product of alphabet size and rollout length. We match this alphabet dependence with a time-invariant shared-address family after two steps, leaving the base and trajectory dimensions unchanged. Its intermediate token identifies a block address, whereas its endpoint reveals one requested bit. The construction yields exact deterministic mistake bounds and realizable learning rates in the probably approximately correct (PAC) model under both feedback protocols. Numerical audits reproduce the feedback gap and its staircase in routing-address capacity.

1 INTRODUCTION

An autoregressive predictor is one next-token rule applied to the context it has generated so far (Bengio et al., 2003; Vaswani et al., 2017; Brown et al., 2020). In endpoint prediction, the learner is evaluated only on the final token, although intermediate tokens may encode routing choices or partial computations absent from that output (Li et al., 2024). This creates two feedback protocols: after predicting, the learner observes the final token alone or the full trajectory of that same rollout, and nothing else differs between the two (Joshi et al., 2025; Doron-Arad et al., 2026). Intermediate tokens can help avoid later mistakes by distinguishing rules that agree on the current endpoint but differ on future prompts.

Existing analyses isolate the effect of generation length most sharply for binary alphabets (Hanneke et al., 2026; Doron-Arad et al., 2026). Longer rollouts can hide more intermediate decisions from final-token feedback, whereas trajectory feedback avoids this growth. A binary token, however, can distinguish only two routing choices, so the binary analyses cannot separate horizon growth from the cost of maintaining a larger internal state. These results therefore leave open whether a larger alphabet, by representing more internal states in one hidden token, can create an additional online-learning cost. In the PAC model, known finite-alphabet bounds contain alphabet-size-dependent terms (Joshi et al., 2025), but they do not establish that such dependence is necessary when the base class is measured by multiclass Littlestone dimension.

We study the feedback gap in deterministic realizable online complexity with the prompts, prediction target, and zero–one loss fixed. Our shared-address construction separates hidden routing from final readout. Its first token identifies an address in a bank shared across parameter blocks, whereas the endpoint reveals only the coordinate requested by the prompt. Figure 1 illustrates how hiding this address can increase the optimal mistake bound under endpoint feedback. The resulting alphabet gap appears after two steps, independently of long rollouts. For the same class and prediction task, full-trajectory feedback reveals the address and removes this gap. Our universal upper bound from

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.

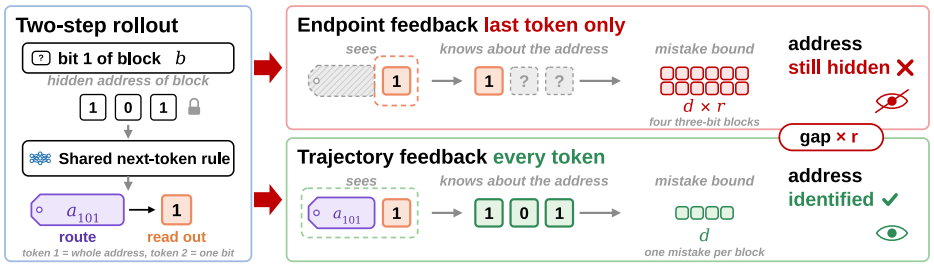


Figure 1: **Hidden routing and the alphabet gap.** After prediction, trajectory feedback reveals the queried block's address, while endpoint feedback reveals only the requested bit. Our shared-address construction yields a mistake gap that grows with alphabet size at fixed generation length, prediction target, and loss.

Section 3 matches the alphabet dependence of the construction, making the two-step alphabet bound tight. We give upper and lower bounds that match on the alphabet dependence, and derive the exact online and PAC consequences of the witness. Joint tightness in alphabet size and rollout length follows from the thresholded binary construction in Appendix F; computational audits reproduce the predicted gap and staircase.

Our contributions are as follows:

1. **General finite-alphabet upper bound.** We show that autoregressive composition can increase final-token online complexity by at most a logarithmic factor in alphabet size and generation length. The argument sharpens the binary analysis through a multiclass sequential-growth bound.
2. **Two-step shared-address witness.** We construct a witness in which one hidden routing token stores an address, while the endpoint reveals only a requested bit. The construction matches the alphabet dependence while keeping both the next-token class and the trajectory-observed class at the prescribed complexity.
3. **Exact learning consequences.** We determine, on the same final-token task and zero–one loss, the optimal deterministic mistake bounds and distribution-free realizable PAC sample complexities under both feedback protocols, and give direct learners that attain the upper bounds.
4. **Computational audits.** We implement the shared-address family and audit its predicted feedback gap and routing-address staircase. Exact searches and online replays reproduce the transition at the routing thresholds; proof-guided PAC calculations recover the predicted dependence on alphabet size, and finite-state consistency tests instantiate the construction.

Section 3 gives the general upper bound. Section 4 develops the shared-address construction, establishes its exact dimensions, and derives mistake bounds and PAC rates. Section 5 checks the dimension identities, feedback gap, and routing-address staircase.

2 Related Work

Two questions organize the work this paper builds on: which class a next-token rule induces, and what the learner observes after predicting. Autoregressive-composition work defines the endpoint and trajectory classes induced by a shared next-token rule, and Transformer-level analyses connect intermediate tokens to expressivity and sample complexity. Feedback-granularity work compares endpoint with trajectory supervision, while sequential-complexity results supply the Littlestone and Natarajan dimensions in which our bounds are stated.

Autoregressive composition. Iterating a shared next-token rule induces both prompt-to-endpoint and prompt-to-trajectory classes. Joshi et al. (2025) analyze their PAC learnability for finite alphabets, Hanneke et al. (2026) give a binary-alphabet PAC taxonomy, and Doron-Arad et al. (2026) develop the corresponding online theory. These works make induced classes, rather than individual next-token decisions, the objects of feedback-dependent learning.

Transformer-level analyses. Li et al. (2024) show how intermediate tokens expand the circuit computations realized by constant-depth, finite-precision Transformers. Yang et al. (2026) derive

architecture-specific sample-complexity bounds for chain-of-thought learning. Both lines tie intermediate generation to a specified architecture and therefore leave open whether the alphabet itself is costly. In the setting below, the next-token class and rollout stay fixed and only feedback changes.

Feedback granularity. Scratchpads and chain-of-thought (CoT) prompting expose intermediate computation (Nye et al., 2021; Wei et al., 2022). In theory, full-trajectory observation removes the horizon dependence in binary-alphabet PAC learning (Hanneke et al., 2026). Doron-Arad et al. (2026, Theorem 2.1) bound the trajectory Littlestone dimension by the base dimension for arbitrary alphabets. Altabaa et al. (2025) characterize the statistical value of CoT supervision through an information measure. Li (2026) studies exact-trace prediction over arbitrary alphabets, where the rollout itself is the target. Here the final token remains the target, and the protocol changes only the revealed part of the trajectory.

Sequential complexity and feedback. Multiclass Littlestone dimension gives the optimal deterministic realizable mistake bound, and its finiteness also characterizes agnostic full-information learnability, including for unbounded label spaces (Littlestone, 1988; Daniely et al., 2015; Hanneke et al., 2023). Both protocols reveal the correct final label, so endpoint feedback remains full information at the prediction target. Bandit multiclass feedback reveals only whether the prediction was correct (Raman et al., 2024); the remaining difficulty is the hidden intermediate trajectory.

The missing gap and our motivation. Intermediate-generation work explains how extra tokens affect expressiveness and sample complexity (Li et al., 2024; Yang et al., 2026), and trajectory-supervision work quantifies the value of observing those tokens (Hanneke et al., 2026; Altabaa et al., 2025). Binary online analyses, however, track generation length rather than alphabet size (Doron-Arad et al., 2026), and finite-alphabet PAC analyses include alphabet terms without showing that the terms are necessary (Joshi et al., 2025). What remains open is whether a larger alphabet alone can raise the final-token Littlestone dimension while the base and trajectory-observed classes keep the base dimension. We settle this with a universal finite-alphabet upper bound and a two-step shared-address family attaining the alphabet dependence; the family also yields exact mistake bounds and realizable PAC rates under both feedback protocols.

3 Feedback Protocols and a General Bound

We compare the endpoint and trajectory classes under a fixed rollout, prediction target, and loss. Expanding endpoint queries into next-token decisions then bounds their complexity through multiclass path counting.

3.1 Autoregressive Rollouts and Feedback

One next-token generator and one recursive rollout induce both observation protocols; the definitions below fix that shared rollout and distinguish the observed label from the predicted target.

For a positive integer m, write $[m] := \{1, \dots, m\}$. Fix a generation length $M \in \mathbb{N}$ with $M \geq 1$. Let Σ be a finite token alphabet of cardinality $K \geq 2$, and let Σ^* be the free monoid of finite words over Σ, including the empty word ϵ. For $x \in \Sigma^*$ and $a \in \Sigma$, write xa for their concatenation. A time-invariant next-token generator is a function $f : \Sigma^* \to \Sigma$ used at every generation step. Time invariance means that the generator may react to its growing context, but the rule itself does not change from one position in the rollout to the next.

For a prompt x, define the generated trajectory recursively by

$$z_0 = x, \qquad y_t = f(z_{t-1}), \qquad z_t = z_{t-1} y_t \quad (t = 1, \dots, M). \tag{1}$$

The end-to-end map $f^{\mathsf{e2e},M}$ returns the final token y_M, while the trajectory map $f^{\mathsf{CoT},M}$ returns $(y_1, \dots, y_M)$; CoT denotes chain of thought. For a *base class* $\mathcal{F} \subseteq \Sigma^{\Sigma^*}$, let $\mathcal{F}^{\mathsf{e2e},M}$ and $\mathcal{F}^{\mathsf{CoT},M}$ be the corresponding classes obtained by varying $f \in \mathcal{F}$. Appendix A gives the explicit maps (6) and class definitions. Throughout, online complexity means the optimal deterministic realizable mistake bound. In both games the learner chooses a token in Σ, predicts y_M, and incurs the corresponding zero–one loss. End-to-end feedback reveals only y_M; CoT feedback reveals $(y_1, \dots, y_M)$. Exact-trace learning instead treats the full trajectory as a single label; it is a different prediction problem from the final-token task studied here.

3.2 Sequential Complexity and the Worst-Case Envelope

The induced endpoint and trajectory classes have different label spaces, so they are compared through combinatorial dimensions defined for arbitrary labels. Littlestone dimension controls deterministic online mistakes, while Natarajan dimension supplies the batch lower bounds used later.

Definition 3.1 (Multiclass Littlestone dimension)**.** For sets $\mathcal{X}, \mathcal{Y}$, let $\mathcal{H} \subseteq \mathcal{Y}^{\mathcal{X}}$ be nonempty. A complete binary tree is Littlestone-shattered by $\mathcal{H}$ if each internal node is labeled by an instance in $\mathcal{X}$, the two outgoing edges at every node have distinct labels in $\mathcal{Y}$, and every root-to-leaf sequence of edge labels is realized by some $h \in \mathcal{H}$ on the queried instances along that path. The largest shattered depth is $\mathrm{Ldim}(\mathcal{H})$, with value 0 if no depth-one tree is shattered.

This is the binary-tree multiclass notion that characterizes deterministic realizable online mistake complexity (Daniely et al., 2015). For an input z and label c, define the label fiber $\mathcal{H}_{z=c} := \{h \in \mathcal{H} : h(z) = c\}$. If a depth-$(q+1)$ tree is shattered with root z and outgoing labels $c_0 \neq c_1$, then, for each $i \in \{0, 1\}$, the nonempty fiber $\mathcal{H}_{z=c_i}$ shatters its depth-q child tree. This root-fiber observation underlies several arguments below.

For batch lower bounds, the argument also uses the multiclass analogue of Vapnik–Chervonenkis (VC) dimension (Natarajan, 1989).

Definition 3.2 (Natarajan dimension)**.** A set $\{x_1, \ldots, x_q\}$ is Natarajan-shattered by $\mathcal{H} \subseteq \mathcal{Y}^{\mathcal{X}}$ if there are labels $c_i^0 \neq c_i^1$ for $i \in [q]$ such that, for every $\sigma \in \{0,1\}^q$, some $h_\sigma \in \mathcal{H}$ satisfies $h_\sigma(x_i) = c_i^{\sigma_i}$ for all $i \in [q]$. The largest such q is $\mathrm{Ndim}(\mathcal{H})$.

Querying a Natarajan-shattered set in a fixed order produces a Littlestone tree, so $\mathrm{Ndim}(\mathcal{H}) \leq \mathrm{Ldim}(\mathcal{H})$. One fixed set of binary choices can therefore certify lower bounds for both dimensions.

The upper-bound argument also uses a complete $\mathcal{Y}$-ary instance tree: every internal node is labeled by an instance and has one outgoing edge for each $c \in \mathcal{Y}$. A root-to-leaf path is $\mathcal{H}$-realizable if one $h \in \mathcal{H}$ agrees with every edge label on that path. Denote the set of realizable paths in a tree $\mathcal{T}$ by $\Pi_{\mathcal{H}}(\mathcal{T})$. This full-branching tree is a path-counting device; Ldim itself remains the binary-tree quantity in Definition 3.1.

For a finite $X_0 \subseteq \mathcal{X}$, write $\mathcal{H}|X_0 := \{h|X_0 : h \in \mathcal{H}\}$. If this restriction is the complete binary cube, then its Natarajan dimension is $|X_0|$, and querying the points in a fixed order gives the same Littlestone lower bound. We also use the elementary finite-class upper bound: a depth-q shattered tree needs 2^q distinct functions, so $\mathrm{Ldim}(\mathcal{H}) \leq \lfloor \log_2 |\mathcal{H}| \rfloor$. Appendix B records the latter inequality where it enters the matching proof.

The preceding dimensions evaluate a fixed class. For fixed M, K, and $d \in \mathbb{Z}_{\geq 0}$, let $\mathfrak{E}_M(d, K)$ denote the supremal endpoint Littlestone dimension over all nonempty base classes on a K-token alphabet whose Littlestone dimension is at most d; Appendix A records the formal supremum. Any two K-token alphabets are equivalent up to relabeling. This envelope allows the universal and constructive results to be compared on the same scale: adaptive path counting controls it from above, while the shared-address family realizes its alphabet dependence at $M = 2$.

3.3 A General Bound on Final-Token Complexity

To bound final-token shattering, expand each endpoint query into its next-token calls and count the resulting realizable paths of the base class.

The proof uses label-fiber dimension descent and Sauer–Shelah-type counting (Sauer, 1972; Shelah, 1972). We extend the binary path-count argument of Doron-Arad et al. (2026) to finite alphabets. Retaining the binomial estimate keeps the alphabet parameter and removes the extra $\log d$ when $K = 2$.

Appendix A states the exact degree-d binomial path count and proves the multiclass sequential-growth lemma. Its summands choose the informative levels and one of $K - 1$ alternatives at each such level.

Theorem 3.3 (Finite-Alphabet End-to-End Bound)**.** *Let $K \geq 2$, $M \geq 1$, and let $\emptyset \neq \mathcal{F} \subseteq \Sigma^{\Sigma^*}$ with $|\Sigma| = K$ and $d = \mathrm{Ldim}(\mathcal{F}) < \infty$. Then $m = \mathrm{Ldim}(\mathcal{F}^{\mathsf{e2e},M})$ is finite. If $d = 0$, then $m = 0$; if $d \geq 1$, then*

$$m < 2d \log_2\big(2e(K-1)M\big). \tag{2}$$

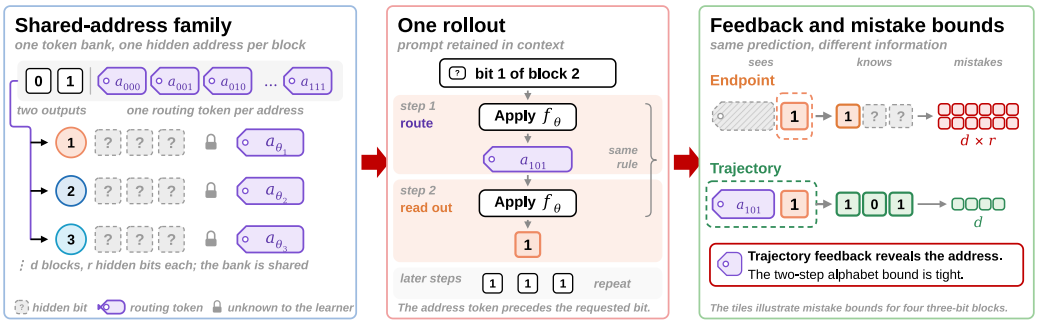


Figure 2: **Shared-address construction.** The d blocks share two output bits and a bank of routing tokens, one for each possible address. On a block query, the same rule f_θ emits the routing token and then the requested bit, using the prompt at both steps. Endpoint feedback reveals only the bit; trajectory feedback also reveals the address. The optimal mistake bounds are dr and d, respectively. At two steps, the endpoint bound matches the alphabet dependence of the general upper bound.

In particular, $\mathrm{Ldim}(\mathcal{F}^{\mathrm{e2e},M}) = O(d \log(KM))$*, with a universal constant.*

Proof sketch. Replace each query in a depth-q shattered end-to-end tree by the M next-token decisions that generate its endpoint, completing non-endpoint branches with a fixed default input that never reconnects. The 2^q endpoint paths then inject into the realizable paths of the resulting depth-qM, K-ary tree. The exact path-count inequality in Appendix A, followed by the standard binomial estimate for $q \geq d \geq 1$, bounds their number by $(eqM(K-1)/d)^d$. Solving the resulting inequality for q yields (2); the case $q < d$ is immediate. Appendix A gives the complete tree expansion and constant calculation.

Equation (2) gives a uniform closed form over all d, K, and M: neither the rollout length nor the alphabet size can increase final-token complexity by more than a logarithmic factor. The shared-address construction below shows that the alphabet term already occurs at $M = 2$ and persists for every $M \geq 2$.

4 Tight Alphabet Separation and Learning Bounds

The shared-address family below realizes the alphabet term in the general bound after two generation steps. Its dimension identities yield exact deterministic mistake bounds and distribution-free realizable PAC sample complexities for final-token prediction under both feedback protocols. Figure 2 traces its routing and readout mechanism and compares the resulting separation with the general upper bound.

4.1 The Shared-Address Construction

Each prompt names a block-coordinate pair. The first output identifies the block address; subsequent outputs return its requested bit. The larger alphabet is used only for routing.

Fix integers $d \geq 1$ and $K \geq 4$. After reserving the two binary output symbols, the largest power-of-two bank of routing tokens stores

$$r := \lfloor \log_2(K-2) \rfloor \quad \text{bits, with} \quad B := \{0,1\}^r, \qquad |B| = 2^r.$$

Thus $r \geq 1$, $2 \leq |B| \leq K-2$, and $dr \geq 1$. For $K = 2, 3$, no routing token remains, so the mechanism starts at $K = 4$. For every address $s = (s_1, \ldots, s_r) \in B$, introduce one routing token a_s. The same $|B|$ routing tokens are reused across all d blocks. The alphabet is the disjoint union of the two binary symbols, the address-indexed tokens, and $K - 2 - |B|$ unused fillers, so $|\Sigma| = K$. Increasing K at fixed r leaves the class and dimensions unchanged; the lower bound tracks $|B|$.

Next, choose dr distinct equal-length prompts $x_{b,j}$, indexed by $b \in [d]$ and $j \in [r]$, all ending in 1. Such a collision-free choice exists with length $L = 1 + \lceil \log_K(dr) \rceil$. Let $X_0 := \{x_{b,j} : b \in [d], j \in [r]\}$. Appendix B.1 records the formal length and canonical lexicographic encoding, and verifies separation from every routed context.

With these prompts fixed, let $\mathcal{R}_{b,j,s}$ denote the persistent routed ray obtained from $x_{b,j}$ by appending the address token a_s and then any number of copies of the selected bit s_j. Appendix B.1 records the exact set representation and its uniqueness.

For a parameter vector $\theta = (\theta_1, \ldots, \theta_d) \in B^d$ and $w \in \Sigma^*$, define the time-invariant next-token generator

$$f_\theta(w) = \begin{cases} a_{\theta_b}, & w = x_{b,j}, \\ s_j, & w \in \mathcal{R}_{b,j,s} \text{ and } \theta_b = s, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

The disjointness properties above make f_θ a well-defined total function on Σ^*. Varying the d block addresses independently gives the base class $\mathcal{F} := \{f_\theta : \theta \in B^d\}$. Distinct parameter vectors disagree on the routing token of any block where their addresses differ. Hence $|\mathcal{F}| = |B|^d = 2^{dr}$. The class therefore contains dr independent address bits, although a single next-token label can identify an entire block address.

On a routed ray whose address satisfies $s = \theta_b$, each generation step appends and re-emits the selected bit. Starting at $x_{b,j}$, the first output is a_{θ_b}, the second is $(\theta_b)_j$, and every subsequent output repeats that bit. The routing tokens can be shared across blocks because the prompt remains part of every routed context and uniquely identifies b, j.

Appendix D gives a finite-state implementation, and Appendix G works out the case $d = 1$, $K = 6$. The product structure yields the exact dimension identities.

4.2 Exact Dimensions under Both Feedback Protocols

At a prompt, base and trajectory labels identify a block address, whereas endpoint labels reveal one coordinate. The next result quantifies the cost of this difference.

Theorem 4.1 (Tight Alphabet Gap). *For every $d \geq 1$ and $K \geq 4$, let $r = \lfloor \log_2(K-2) \rfloor$. The alphabet, prompt set, and finite class constructed above satisfy*

1. *$|\mathcal{F}| = 2^{dr}$ and $\mathrm{Ldim}(\mathcal{F}) = \mathrm{Ndim}(\mathcal{F}) = d$;*
2. *for every $M \geq 2$, $\mathcal{F}^{\mathsf{e2e},M} \subseteq \{0,1\}^{\Sigma^*}$, and*
$$(\mathcal{F}^{\mathsf{e2e},M})\restriction X_0 = \{0,1\}^{X_0}, \qquad \mathrm{Ldim}(\mathcal{F}^{\mathsf{e2e},M}) = \mathrm{Ndim}(\mathcal{F}^{\mathsf{e2e},M}) = dr;$$
3. *for every $M \geq 1$, $\mathrm{Ldim}(\mathcal{F}^{\mathsf{CoT},M}) = \mathrm{Ndim}(\mathcal{F}^{\mathsf{CoT},M}) = d$.*

At $M = 1$, the end-to-end class equals $\mathcal{F}$ and also has both dimensions equal to d.

Starting from $x_{b,j}$, the routing token selects θ_b, and the next output reads its j-th bit. Hence, for every $M \geq 2$,

$$f_\theta^{\mathsf{e2e},M}(x_{b,j}) = (\theta_b)_j. \tag{4}$$

Proof sketch of Theorem 4.1. Every genuine base- or trajectory-label split has a branch that fixes one previously unresolved block address; induction gives their upper bounds. Independent choices in one prompt per block give both lower bounds. Equation (4) turns X_0 into a complete dr-point binary cube. The class cardinality supplies the matching endpoint upper bound, while the one-step case follows directly from the rollout recursion. Appendix B states the product-space lemmas and gives the complete fiber argument for arbitrary initial contexts.

4.3 Matching Worst-Case Bounds

The shared-address family gives an alphabet lower bound for the worst-case envelope. The general upper bound matches its alphabet dependence at two steps.

Corollary 4.2. *For every $d \geq 1$, $K \geq 4$, and $M \geq 2$,*

$$d\lfloor \log_2(K-2) \rfloor \leq \mathfrak{E}_M(d,K) < 2d \log_2\big(2e(K-1)M\big). \tag{5}$$

In particular, $\mathfrak{E}_2(d,K) = \Theta(d \log K)$, with uniform constants over the full alphabet range.

Proof sketch. Theorems 4.1 and 3.3 give (5); setting $M = 2$ gives the two-step rate. Appendix F verifies the constants.

Appendix F combines the alphabet witness with the binary horizon construction to obtain the corresponding $\Theta(d \log(KM))$ envelope in the large-horizon regime.

4.4 Online Mistake Bounds

Since multiclass Littlestone dimension equals the optimal deterministic realizable mistake bound, the exact identities above determine the online cost of hiding the trajectory. Fix $f_\star \in \mathcal{F}$. Both protocols predict its final token under zero–one loss; after prediction, they reveal either the endpoint or the full trajectory. The trajectory is used only to update the learner state. Appendix C.1 gives the per-round notation.

Corollary 4.3. *For the class in Theorem 4.1 and every $M \geq 2$, let $D := dr = d\lfloor \log_2(K-2) \rfloor$. The optimal deterministic realizable final-token mistake bound is D under end-to-end feedback and d under CoT feedback.*

Proof sketch. The end-to-end statement follows from $\mathrm{Ldim}(\mathcal{F}^{\mathsf{e2e},M}) = D$ and the multiclass Littlestone characterization (Littlestone, 1988; Daniely et al., 2015). Under CoT feedback, a learner may run an optimal realizable learner for $\mathcal{F}^{\mathsf{CoT},M}$ internally, as an analysis device, and report the last coordinate of its predicted trajectory. Every final-token mistake is then a trajectory mistake, so Theorem 4.1 gives the upper bound d. For the matching lower bound, a blockwise query sequence forces one mistake per block while remaining consistent with a single fixed target $f_\star$; Appendix C gives the full construction.

The mistake-count ratio is r. On a block's first prompt, final feedback reveals one bit and trajectory feedback all r. The mistake bounds are information-theoretic; Table 3 lists the direct learners used in the audits and their resource costs for dr-bit targets.

4.5 Realizable PAC Separation

The same feedback distinction also changes distribution-free batch complexity. For a distribution P on Σ^*, an n-sample learner receives independent $x_i \sim P$ together with either $f_\star^{\mathsf{e2e},M}(x_i)$ or $f_\star^{\mathsf{CoT},M}(x_i)$, and returns an endpoint predictor $g : \Sigma^* \to \Sigma$. In both protocols, its risk is $P(g(x) \neq f_\star^{\mathsf{e2e},M}(x))$. Let $n_{\mathrm{PAC}}^{\mathsf{e2e}}(\varepsilon, \delta)$ and $n_{\mathrm{PAC}}^{\mathsf{CoT}}(\varepsilon, \delta)$ denote the sample sizes needed for error at most ε with probability at least $1 - \delta$.

Corollary 4.4 (PAC Feedback Separation)**.** *For the class in Theorem 4.1, $M \geq 2$, and $D = dr$, the distribution-free realizable PAC sample complexities satisfy*

$$n_{\mathrm{PAC}}^{\mathsf{e2e}}(\varepsilon, \delta) = \Theta\left(\frac{D + \ln(1/\delta)}{\varepsilon}\right), \qquad n_{\mathrm{PAC}}^{\mathsf{CoT}}(\varepsilon, \delta) = \Theta\left(\frac{d + \ln(1/\delta)}{\varepsilon}\right)$$

for $0 < \varepsilon \leq 1/8$ and $0 < \delta \leq 1/100$, even when improper learners are allowed.

Proof sketch. The end-to-end binary cube gives the lower bound, and a union bound over the 2^D induced maps gives the matching upper bound. The CoT learner uses this block structure: a positive example from an unknown block identifies its entire address, and a union bound over the 2^d sets of unseen blocks gives the upper bound. Restricting each block to two addresses that differ in one readout bit gives the matching lower bound. Appendix C gives the complete proof. Both rates count labeled examples; fixed-width feedback bits per sample are tabulated separately in Table 3.

Both protocols share the same confidence term; the gain comes from the information each sample carries. The hidden address is therefore the single source of both gaps, online and batch, under a fixed prediction task and loss.

5 Experiments

We audit the shared-address construction with exact dimension searches and deterministic version-space replays. Exact search covers the finite classes, while factorized updates extend the alphabet and horizon sweeps. All instances follow Section 4.1; Appendix E gives the methods and parameter ranges.

5.1 Exact Dimension Computation

The exact-search dimensions of the tested base, final-token, and trajectory classes agree with Theorem 4.1. Table 1 shows the informative boundary cases. One-step generation leaves the endpoint equal to the base prediction. At two steps, the endpoint dimension expands once an address contains more than one bit, while the base and trajectory dimensions remain fixed. The one-bit $K = 4$ control has no gap.

Table 1: Representative exact-search dimensions. $D = dr$ is the two-step endpoint dimension, and $\Delta_{\mathrm{F-T}}$ is final minus trajectory dimension for both Ldim and Ndim.

Instance					Base		Final		Trajectory		Gap
d	K	M	r	$\lvert\mathcal{F}\rvert$	$D = dr$	Ldim	Ndim	Ldim	Ndim	Ldim	Ndim	$\Delta_{\mathrm{F-T}}$
Controls												
3	4	2	1	8	3	3	3	3	3	3	3	0
1	6	1	2	4	2	1	1	1	1	1	1	0
Two-step separation												
1	6	2	2	4	2	1	1	2	2	1	1	1
2	6	2	2	16	4	2	2	4	4	2	2	2
3	8	2	2	64	6	3	3	6	6	3	3	3

The finite input representation uses hypothesis-evaluation signatures; Appendix E describes the product-machine reduction.

5.2 Online Mistake Counts

We compare the Standard Optimal Algorithm (SOA) with majority voting over the remaining hypotheses' endpoint predictions. Table 2 summarizes twelve adversarial runs varying the learner, tie rule, and feedback protocol.

Table 2: Feedback and learner comparison at $d = 3, K = 8, M = 2$. SOA and Majority entries give cumulative mistakes under the 0-first/1-first tie rules after six adversarial rounds. All twelve transcripts are replayed against their fixed targets; Theory gives the optimal mistake bound.

Feedback information			**Endpoint mistakes**		
Protocol	**Feedback**	**Resolved unit**	**SOA**	**Majority**	**Theory**
Final	Endpoint	One coordinate	6/6	6/6	6
Trajectory	Full trajectory	One block address	3/3	3/3	3
Trace-blind	Endpoint only	One coordinate	6/6	6/6	6

Across these twelve runs, the counts do not change with learner or tie rule. Trace-blind receives the same generated trajectory but discards its intermediate tokens; its mistake count matches that of final feedback. A full trajectory fixes one block address rather than one coordinate and halves the number of mistakes here because $r = 2$. Figure 3(a) shows the corresponding SOA dynamics on a fixed target; Majority overlaps it on this symmetric sequence.

Panels (b)–(c) isolate the two parameters in Theorem 4.1. Alphabet size changes the result only when another routing bit fits, whereas generation length activates the readout at the second token. The M-flat curve is witness-specific; worst-case horizon growth comes from the binary construction in Appendix F.

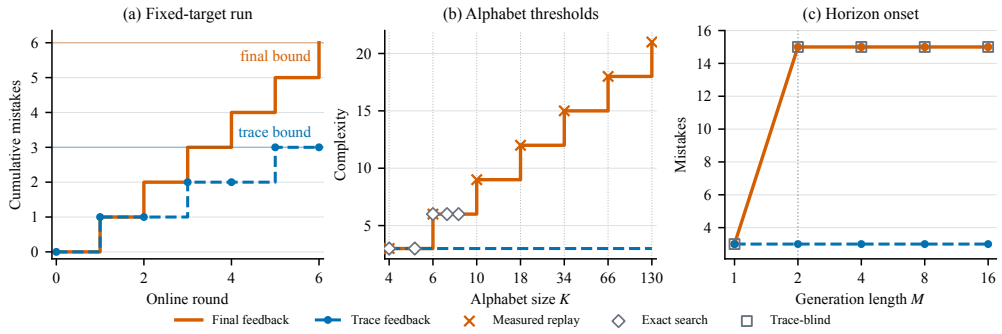


Figure 3: Deterministic feedback gap and parameter controls. **(a)** At $d = 3, K = 8, M = 2$, SOA makes six mistakes with final feedback and three with trajectory feedback on the same target. **(b)** At $d = 3, M = 2$, lines give the theorem staircase, crosses are 0-first replays, and diamonds are exact searches for $K \leq 8$; $K - 2$ is log-spaced. **(c)** At $d = 3, K = 34$, all protocols make three mistakes for $M = 1$; at each tested $M \geq 2$, final and trace-blind make fifteen while trajectory feedback remains at three; the M-flat curve is witness-specific.

Figure 4 separates the PAC dimension and confidence terms. Address capacity shifts the final-feedback threshold but leaves the trajectory threshold fixed; confidence instead moves both curves. Corollary 4.4 has the same decomposition: feedback changes the effective number of learnable units, and the confidence cost is common to both protocols.

On small classes, direct empirical risk minimization (ERM) enumeration agrees with the dynamic program's unseen-mass risks. The tested Moore machines agree with the direct generator on every checked word and use one transition per input token. Figure 5 and Table 6 use state minimization to measure the slack in the explicit construction, separately from the output comparison.

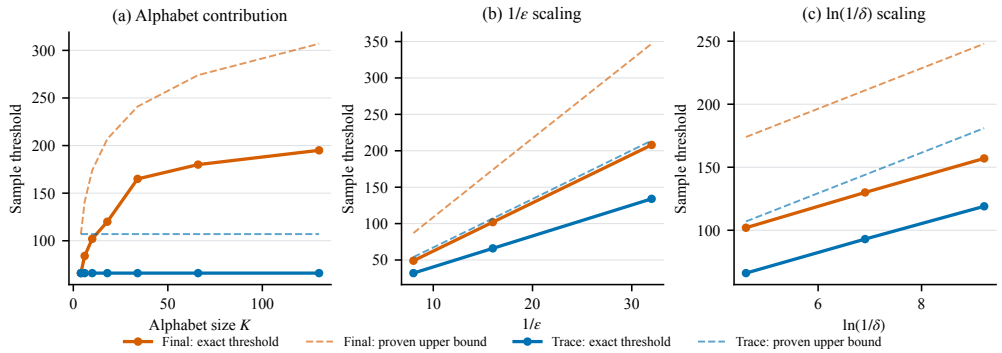


Figure 4: PAC thresholds for the fixed minimum-index ERM. Solid curves give the larger threshold over the protocol-specific dimension and common confidence distributions; dashed curves give distribution-free upper bounds. Panel settings are in Appendix E.2.

6 Conclusion

We characterize how hidden intermediate routing changes deterministic realizable autoregressive learning over unrestricted next-token classes on finite alphabets, with fixed prediction target, loss, and generator class. A multiclass sequential-growth argument bounds the growth of final-token complexity for arbitrary finite alphabets and rollout lengths. Our finite-state shared-address family matches the alphabet dependence after two steps while preserving the original complexity of the next-token and trajectory-observed classes; routing capacity is therefore a source of worst-case difficulty distinct from generation length. The thresholded joint alphabet-and-horizon statement is in Appendix F. The same construction yields matching online and realizable PAC separations. The computational study mirrors this structure: exact searches recover the dimensions, replays reproduce the gap, and Moore-machine checks instantiate the realization. All three point to information granularity: trajectory feedback reveals an address, endpoint feedback a coordinate, so the separation tracks the information revealed.

AI USE STATEMENT

Generative AI tools assisted language editing and formatting.

REFERENCES

Awni Altabaa, Omar Montasser, and John Lafferty. CoT information: Improved sample complexity under chain-of-thought supervision. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 24822–24862. Curran Associates, Inc., 2025. doi: 10.52202/085713-0834.

Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Jauvin. A neural probabilistic language model. *Journal of Machine Learning Research*, 3:1137–1155, 2003. URL https://www.jmlr.org/papers/v3/bengio03a.html.

Tom B. Brown et al. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pp. 1877–1901, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html.

Amit Daniely, Sivan Sabato, Shai Ben-David, and Shai Shalev-Shwartz. Multiclass learnability and the ERM principle. *Journal of Machine Learning Research*, 16(72):2377–2404, 2015. URL https://jmlr.org/papers/v16/daniely15a.html.

Ilan Doron-Arad, Idan Mehalel, and Elchanan Mossel. A theory of online learning with autoregressive chain-of-thought reasoning. *arXiv preprint arXiv:2605.06819*, 2026. doi: 10.48550/arXiv.2605.06819. URL https://arxiv.org/abs/2605.06819.

Andrzej Ehrenfeucht, David Haussler, Michael Kearns, and Leslie Valiant. A general lower bound on the number of examples needed for learning. *Information and Computation*, 82(3):247–261, 1989. doi: 10.1016/0890-5401(89)90002-3.

Steve Hanneke. The optimal sample complexity of PAC learning. *Journal of Machine Learning Research*, 17(38):1–15, 2016. URL https://jmlr.org/papers/v17/15-389.html.

Steve Hanneke, Shay Moran, Vinod Raman, Unique Subedi, and Ambuj Tewari. Multiclass online learning and uniform convergence. In *Proceedings of the Thirty-Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pp. 5682–5696, 2023. URL https://proceedings.mlr.press/v195/hanneke23b.html.

Steve Hanneke, Idan Mehalel, and Shay Moran. Sample complexity of autoregressive reasoning: Chain-of-thought vs. end-to-end. *arXiv preprint arXiv:2604.12013*, 2026. doi: 10.48550/arXiv.2604.12013. URL https://arxiv.org/abs/2604.12013.

Nirmit Joshi, Gal Vardi, Adam Block, Surbhi Goel, Zhiyuan Li, Theodor Misiakiewicz, and Nathan Srebro. A theory of learning with autoregressive chain of thought. In Nika Haghtalab and Ankur Moitra (eds.), *Proceedings of Thirty Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pp. 3161–3212. PMLR, 2025. URL https://proceedings.mlr.press/v291/joshi25a.html.

Zhiyuan Li. The optimal sample complexity of learning autoregressive chain-of-thought. *arXiv preprint arXiv:2607.07423*, 2026. doi: 10.48550/arXiv.2607.07423. URL https://arxiv.org/abs/2607.07423.

Zhiyuan Li, Hong Liu, Denny Zhou, and Tengyu Ma. Chain of thought empowers transformers to solve inherently serial problems. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 11911–11943, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/3309b4112c9f04a993f2bbdd0274bba1-Abstract-Conference.html.

Nick Littlestone. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Machine Learning*, 2(4):285–318, 1988. doi: 10.1007/BF00116827.

B. K. Natarajan. On learning sets and functions. *Machine Learning*, 4:67–97, 1989. doi: 10.1007/BF00114804.

Maxwell Nye, Anders Johan Andreassen, Guy Gur-Ari, Henryk Michalewski, Jacob Austin, David Bieber, David Dohan, Aitor Lewkowycz, Maarten Bosma, David Luan, Charles Sutton, and Augustus Odena. Show your work: Scratchpads for intermediate computation with language models. *arXiv preprint arXiv:2112.00114*, 2021. doi: 10.48550/arXiv.2112.00114. URL https://arxiv.org/abs/2112.00114.

Ananth Raman, Vinod Raman, Unique Subedi, Idan Mehalel, and Ambuj Tewari. Multiclass online learnability under bandit feedback. In *Proceedings of the Thirty-Fifth International Conference on Algorithmic Learning Theory*, volume 237 of *Proceedings of Machine Learning Research*, pp. 997–1012, 2024. URL https://proceedings.mlr.press/v237/raman24a.html.

Norbert Sauer. On the density of families of sets. *Journal of Combinatorial Theory, Series A*, 13(1): 145–147, 1972. doi: 10.1016/0097-3165(72)90019-2.

Saharon Shelah. A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific Journal of Mathematics*, 41(1):247–261, 1972. doi: 10.2140/pjm.1972.41.247.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, pp. 5998–6008, 2017. URL https://papers.nips.cc/paper/7181-attention-is-all-you-need.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 24824–24837, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/9d5609613524ecf4f15af0f7b31abca4-Abstract-Conference.html.

Chenxiao Yang, Nathan Srebro, and Zhiyuan Li. Tight sample complexity of transformers. In Steve Hanneke and Tor Lattimore (eds.), *Proceedings of Thirty Ninth Conference on Learning Theory*, volume 336 of *Proceedings of Machine Learning Research*, pp. 6887–6923. PMLR, 2026. URL https://proceedings.mlr.press/v336/yang26a.html.

A Proofs for the General Upper Bound

Fix a K-token alphabet Σ and a rollout length $M \geq 1$. For the sequence in (1), the endpoint and trajectory maps are

$$f^{\mathsf{e2e},M}(x) := y_M, \qquad f^{\mathsf{CoT},M}(x) := (y_1, \ldots, y_M). \tag{6}$$

Thus $f^{\mathsf{e2e},M} : \Sigma^* \to \Sigma$ and $f^{\mathsf{CoT},M} : \Sigma^* \to \Sigma^M$. For $M = 2$, in particular, $f^{\mathsf{e2e},2}(x) = f(xf(x))$. For a base class $\mathcal{F} \subseteq \Sigma^{\Sigma^*}$, the induced classes are $\mathcal{F}^{\mathsf{e2e},M} := \{f^{\mathsf{e2e},M} : f \in \mathcal{F}\}$ and $\mathcal{F}^{\mathsf{CoT},M} := \{f^{\mathsf{CoT},M} : f \in \mathcal{F}\}$. The worst-case end-to-end envelope is

$$\mathfrak{E}_M(d, K) := \sup \left\{ \mathrm{Ldim}(\mathcal{G}^{\mathsf{e2e},M}) : \emptyset \neq \mathcal{G} \subseteq \Sigma^{\Sigma^*},\ \mathrm{Ldim}(\mathcal{G}) \leq d \right\}.$$

A.1 Sequential Growth

For integers $n, d \geq 0$, define the exact multiclass path count

$$S_K(n, d) := \sum_{i=0}^{\min\{d,n\}} \binom{n}{i} (K-1)^i.$$

The index i selects the informative levels, and the factor $(K-1)^i$ selects their nonreference labels.

Lemma A.1 (Multiclass sequential growth)**.** *Let $\mathcal{H} \subseteq \mathcal{Y}^{\mathcal{X}}$ be nonempty, $2 \leq |\mathcal{Y}| = K < \infty$, and $\mathrm{Ldim}(\mathcal{H}) \leq d$. For every complete $\mathcal{Y}$-ary instance tree $\mathcal{T}$ of depth n,*

$$|\Pi_{\mathcal{H}}(\mathcal{T})| \leq S_K(n, d).$$

Proof of Lemma A.1. We induct simultaneously on $n + d$. The claim is immediate for $n = 0$. If $d = 0$, all functions in $\mathcal{H}$ agree at every instance (otherwise a depth-one tree is shattered), so exactly one path is realizable.

Now let $n, d \geq 1$, let x label the root, and put $\mathcal{H}_c = \mathcal{H}_{x=c}$. At most one nonempty fiber can have Littlestone dimension d: if two fibers had dimension d, attach their respective depth-d shattered trees below a new root x, using their two distinct labels, to obtain a depth-$(d + 1)$ tree for $\mathcal{H}$. Every other nonempty fiber has dimension at most $d - 1$. If a dimension-d fiber exists, designate it; otherwise designate any nonempty fiber. For each label c, let $\mathcal{T}_c$ be the child subtree reached by the edge labeled c. Applying the induction hypothesis gives

$$|\Pi_{\mathcal{H}}(\mathcal{T})| = \sum_{c \in \mathcal{Y}} |\Pi_{\mathcal{H}_c}(\mathcal{T}_c)| \leq S_K(n-1, d) + (K-1)S_K(n-1, d-1).$$

Pascal's identity makes the right-hand side exactly $S_K(n, d)$, including the case $d \geq n$.

A.2 Finite-Alphabet End-to-End Bound

Proof of Theorem 3.3. Fix a tree of finite depth q shattered by the end-to-end class, and denote it by $\mathcal{T}_{\mathrm{e2e}}$. A node v has prompt x_v and two distinct outgoing end-to-end labels c_v^0, c_v^1.

Expand each end-to-end node into M next-token levels. At level t, after labels $\ell_1, \ldots, \ell_{t-1}$, query the prompt-conditioned context $x_v \ell_1 \cdots \ell_{t-1}$. Give that node all K outgoing labels in Σ. At the end of the segment, if $\ell_M = c_v^i$ for $i \in \{0, 1\}$, the path moves to the i-th live child of v; every other final label leads to a dead branch. Label the remaining nodes in that subtree by the fixed input $\bar{x} = \epsilon$ and never reconnect them to a live child. Repeating at every live child produces a complete Σ-ary instance tree $\mathcal{T}_{\mathrm{nt}}$ of depth qM.

Each of the 2^q end-to-end paths has a realizing generator. Recursion (1) determines a unique next-token path whose final label in every live segment is the prescribed end-to-end label, and two end-to-end paths remain distinct at their first divergent segment. Hence

$$2^q \leq |\Pi_{\mathcal{F}}(\mathcal{T}_{\mathrm{nt}})| \leq S_K(qM, d),$$

where the second inequality is Lemma A.1.

If $d = 0$, then $S_K(qM, 0) = 1$, so no positive depth is shattered and $m = 0$. For $d \geq 1$, the claimed bound is automatic when $q < d$. When $q \geq d$, put $n = qM$ and $\alpha = d/(n(K-1))$. Since $0 < \alpha \leq 1$,

$$\begin{aligned} S_K(n, d) &\leq \alpha^{-d} \sum_{i=0}^{n} \binom{n}{i} \big((K-1)\alpha\big)^i \\ &= \alpha^{-d} \big(1 + (K-1)\alpha\big)^n \leq \left(\frac{en(K-1)}{d}\right)^d. \end{aligned}$$

Writing $\rho = q/d$ gives $2^\rho \leq e(K-1)M\rho = A\rho$, where $A = e(K-1)M$. If $1 \leq \rho < 2$, then $\rho < 2\log_2(2A)$. If $\rho \geq 2$, the function $2^\rho/\rho$ is increasing, and at $\rho_0 = 2\log_2(2A)$,

$$\frac{2^{\rho_0}}{\rho_0} = \frac{2A^2}{\log_2(2A)} > A.$$

Therefore $\rho < \rho_0$, proving $m < 2d\log_2(2e(K-1)M)$.

B Proofs for the Shared-Address Separation

The separation proof rests on a blockwise invariant: every genuine split has a branch that fixes a previously unresolved address block. The same invariant governs base labels and complete trajectories. Concretely, a full trajectory identifies one block address at a time, whereas for $M \geq 2$ the final outputs realize a binary cube over all dr block-coordinate pairs; the dimension calculations make this precise.

B.1 Explicit Encoding and Elementary Bounds

Set

$$L := 1 + \lceil \log_K(dr) \rceil. \tag{7}$$

Fix an ordering of Σ, take the first dr length-$(L-1)$ words in lexicographic order, and append the symbol 1. Since $K^{L-1} \geq dr$, this gives distinct prompts $x_{b,j}$ of common length L, all ending in 1. For $b \in [d]$, $j \in [r]$, and $s \in B$, the persistent routed ray is

$$\mathcal{R}_{b,j,s} := \{x_{b,j} a_s (s_j)^q : q \in \mathbb{Z}_{\geq 0}\},$$

where $(s_j)^q$ is the length-q constant word and $(s_j)^0 = \epsilon$. A prompt cannot be a routed-ray word because the latter has length at least $L+1$. If two routed-ray words coincide, their first L symbols determine b, j, the next symbol determines s, and their common length determines the exponent. Thus the prompt and ray cases in (3) are disjoint and every ray representation is unique.

We also use the finite-class inequality

$$\mathrm{Ldim}(\mathcal{H}) \leq \lfloor \log_2 |\mathcal{H}| \rfloor, \tag{8}$$

because a depth-q shattered binary tree requires a distinct function for each of its 2^q paths.

For nonempty $S_b \subseteq B$, define

$$\mathcal{F}(S_1, \ldots, S_d) := \{f_\theta : \theta_b \in S_b \text{ for every } b \in [d]\}, \qquad u(S_1, \ldots, S_d) := \big|\{b \in [d] : |S_b| \geq 2\}\big|.$$

Lemma B.1 (Blockwise dimension)**.** *For all nonempty $S_1, \ldots, S_d \subseteq B$,*

$$\mathrm{Ldim}\big(\mathcal{F}(S_1, \ldots, S_d)\big) = u(S_1, \ldots, S_d).$$

Fix such $S_1, \ldots, S_d$, abbreviate $\mathcal{G} = \mathcal{F}(S_1, \ldots, S_d)$, and let $u = u(S_1, \ldots, S_d)$.

B.2 Base-Class Lower Bound

Each unresolved block contributes one independent level to a shattered tree.

Let $U = \{b : |S_b| \geq 2\}$ and order it as $b_1, \ldots, b_u$. Choose distinct $s^0_{b_t}, s^1_{b_t} \in S_{b_t}$. At every node of level t, query $x_{b_t,1}$ and use edge labels $a_{s^0_{b_t}}, a_{s^1_{b_t}}$. For a path $\sigma \in \{0,1\}^u$, set $\theta_{b_t} = s^{\sigma_t}_{b_t}$; in each resolved block, use its unique allowed address. This parameter belongs to the product and realizes the path. Hence $\mathrm{Ldim}(\mathcal{G}) \geq u$.

B.3 Base-Class Fiber Classification

There are three disjoint input types.

Prompt input. For $w = x_{b,j}$, the possible labels are $\{a_s : s \in S_b\}$, and

$$\mathcal{G}_{x_{b,j} = a_s} = \mathcal{F}(S_1, \ldots, S_{b-1}, \{s\}, S_{b+1}, \ldots, S_d) \quad (s \in S_b). \tag{9}$$

Every other fiber is empty. If $|S_b| = 1$, the query is constant. If $|S_b| \geq 2$, every nonempty fiber has unresolved count $u - 1$. The fact that a_s is shared by all blocks creates no additional fiber: the queried prompt fixes the index b.

One-bit ray input. Suppose $w \in \mathcal{R}_{b,j,s}$ and $s_j = 1$. If $s \notin S_b$, the query is constant 0. If $s \in S_b$, its only possible labels are 1 and 0, with

$$\mathcal{G}_{w=1} = \mathcal{F}(S_1, \ldots, S_{b-1}, \{s\}, S_{b+1}, \ldots, S_d), \tag{10}$$

$$\mathcal{G}_{w=0} = \mathcal{F}(S_1, \ldots, S_{b-1}, S_b \setminus \{s\}, S_{b+1}, \ldots, S_d). \tag{11}$$

The second fiber is empty if $S_b = \{s\}$, so that case is not a split. Whenever both fibers are nonempty, the exceptional fiber resolves block b. The background fiber retains count u only when $|S_b \setminus \{s\}| \geq 2$.

Zero-bit rays and default input. If $w \in \mathcal{R}_{b,j,s}$ and $s_j = 0$, the matching clause and the default clause both output 0, so the query is constant. Every word outside the prompts and routed rays is constant 0 by definition.

These cases are exhaustive and disjoint by the encoding above.

B.4 Base-Class Upper Bound

The fiber classification gives the matching upper bound: every genuine split has a branch with one fewer unresolved block.

For $u = 0$, every coordinate set is a singleton and $\mathcal{G}$ contains one function. Suppose the upper bound holds for every product with at most $u - 1$ unresolved blocks. If a depth-$(u + 1)$ tree were shattered by a product with count u, its root would have two distinct nonempty label fibers, each required to shatter depth u. A prompt root sends both fibers to count $u - 1$. A nonconstant ray root has an exceptional fiber of count at most $u - 1$. The induction hypothesis makes that fiber's dimension at most $u - 1$, a contradiction. This proves the base-class upper bound.

B.5 Persistent Readout

Lemma B.2 (Persistent readout)**.** *For every $\theta \in B^d$, $b \in [d]$, $j \in [r]$, and $M \geq 1$, the prompt-conditioned outputs satisfy $y_1 = a_{\theta_b}$ and $y_t = (\theta_b)_j$ for $2 \leq t \leq M$. Consequently, $f_\theta^{\mathsf{e2e},M}(x_{b,j}) = (\theta_b)_j$ for every $M \geq 2$.*

Proof of Lemma B.2. The prompt clause gives $y_1 = a_{\theta_b}$. Before the second application, the current context is $x_{b,j} a_{\theta_b} \in \mathcal{R}_{b,j,\theta_b}$, with ray exponent zero, so the matching-ray clause gives $y_2 = (\theta_b)_j$. Inductively, before application $t \geq 2$ the context is

$$x_{b,j} a_{\theta_b} \big((\theta_b)_j\big)^{t-2} \in \mathcal{R}_{b,j,\theta_b}.$$

The same clause re-emits $(\theta_b)_j$, proving the induction and the final-token identity.

B.6 CoT Fiber Classification

Complete trajectories induce the same product fibers as their first informative token, preserving the unresolved-block count used above.

Lemma B.3 (Trajectory block dimension)**.** *For every $M \geq 1$ and every nonempty product version space,*

$$\mathrm{Ldim}\big(\mathcal{F}(S_1, \ldots, S_d)^{\mathsf{CoT},M}\big) = \mathrm{Ndim}\big(\mathcal{F}(S_1, \ldots, S_d)^{\mathsf{CoT},M}\big) = u(S_1, \ldots, S_d).$$

Fix $M \geq 1$. At a prompt $x_{b,j}$, parameters with $\theta_b = s$ produce

$$\tau_{j,s}^{(M)} := \begin{cases} (a_s), & M = 1, \\ (a_s, s_j, \ldots, s_j), & M \geq 2. \end{cases} \tag{12}$$

Because the first component is a_s, the fiber of $\tau_{j,s}^{(M)}$ is precisely the product in (9). For $c \in \Sigma$, write $c^M = (c, \ldots, c) \in \Sigma^M$ for the length-M constant trajectory.

At a ray $\mathcal{R}_{b,j,s}$ with $s_j = 1$ and $s \in S_b$, the fibers corresponding to 1^M and 0^M are given by (10) and (11); if $s \notin S_b$, the trajectory is constantly 0^M. On a ray with $s_j = 0$, the trajectory is also always 0^M.

For an initial context outside all designated sets, the first output is 0. After $t \geq 1$ default outputs, the current word has form $w0^t$. Such a word cannot be a prompt, since prompts end in 1. It cannot be a ray of exponent zero, since those words end in a_s, and it cannot be a positive-exponent one-bit ray, since those words end in 1. It cannot be a zero-bit ray either, for then w itself would lie in that ray. Hence its entire trajectory is 0^M.

The CoT-label fibers therefore have the same unresolved-block structure as the base-class fibers. The same induction gives the CoT class Littlestone dimension at most u. Conversely, querying one prompt per unresolved block and choosing two trajectory labels from (12) Natarajan-shatters u points. This proves the two dimension equalities in Lemma B.3.

B.7 Proof of the Tight Alphabet Gap

Proof of Theorem 4.1. For the base class $\mathcal{F}$, each $S_b = B$ has at least two elements. Lemma B.1 gives $\mathrm{Ldim}(\mathcal{F}) = d$. For the base-class Natarajan lower bound, query $x_{b,1}$ for each $b \in [d]$, choose two distinct addresses in every block, and use their routing tokens as labels. Independent choices realize all 2^d labelings, so $\mathrm{Ndim}(\mathcal{F}) = d$.

For $M \geq 2$, a prompt produces a routing token followed by binary readouts, while a routed-ray input produces only binary readouts. Every other initial context produces zeros. The preceding case analysis makes this statement exhaustive. By Lemma B.2,

$$(f_\theta^{\mathsf{e2e},M})(x_{b,j}) = (\theta_b)_j.$$

Given any $\ell \in \{0,1\}^{X_0}$, set $\theta_b = (\ell(x_{b,1}), \ldots, \ell(x_{b,r}))$. The resulting generator realizes ℓ, so the end-to-end restriction is the complete dr-point binary cube and both dimensions are at least dr. Conversely, $|\mathcal{F}^{\mathsf{e2e},M}| \leq |\mathcal{F}| = 2^{dr}$, so (8) gives the matching Littlestone upper bound. Lemma B.3 with $S_b = B$ gives both CoT dimensions as d. Finally, $M = 1$ follows from $\mathcal{F}^{\mathsf{e2e},1} = \mathcal{F}$.

C Proofs for the Learning Consequences

The persistent-readout identity makes the end-to-end restriction a complete binary cube. Together with the block structure, it yields the exact online and PAC feedback comparisons.

Let $D = dr$. For each $M \geq 2$, define

$$R_M : B^d \longrightarrow \{0,1\}^{X_0}, \qquad R_M(\theta)(x_{b,j}) := (f_\theta^{\mathsf{e2e},M})(x_{b,j}).$$

Lemma B.2 gives

$$R_M(\theta)(x_{b,j}) = (\theta_b)_j.$$

The map is injective because all coordinates of every θ_b occur, and it is surjective because each $\ell \in \{0,1\}^{X_0}$ determines $\theta_b = (\ell(x_{b,1}), \ldots, \ell(x_{b,r}))$. Thus R_M is a bijection for every $M \geq 2$, and its image is the complete binary cube of 2^D distinct maps.

C.1 Exact Feedback Mistake Bounds

Fix $f_\star = f_\theta \in \mathcal{F}$, with unique parameter $\theta \in B^d$, and let $\mathbf{1}\{E\}$ denote the indicator of an event E. On round i, the learner receives x_i, predicts $\widehat{y}_i$, and incurs $\mathbf{1}\left\{\widehat{y}_i \neq f_\star^{\mathsf{e2e},M}(x_i)\right\}$. It then observes either the endpoint or the full trajectory.

End-to-end feedback is standard online learning of $\mathcal{F}^{\mathsf{e2e},M}$. Its optimal deterministic realizable mistake bound is its Littlestone dimension (Littlestone, 1988; Daniely et al., 2015), which Theorem 4.1 gives as D.

For the CoT-feedback upper bound, run an optimal realizable online learner for $\mathcal{F}^{\mathsf{CoT},M}$ internally and report the last coordinate of its predicted trajectory. Every end-to-end mistake is also a mistake on the predicted CoT trajectory, so the number of mistakes is at most $\mathrm{Ldim}(\mathcal{F}^{\mathsf{CoT},M}) = d$.

Explicit learners for the witness. Inputs outside the prompt and routed-ray sets have constant endpoint 0. For other inputs, parse the unique block b and maintain its surviving addresses $S_b \subseteq B$. End-to-end feedback predicts the majority endpoint in $\{0,1\}$; every mistake halves S_b, for at most r mistakes per block. Under CoT feedback, predict 0 until a prompt trace or positive routed-ray trace identifies θ_b, for at most one mistake per block.

For the reverse inequality, query $x_{b,1}$ once in each previously unseen block b. Before this query the address θ_b is unrestricted. If the learner predicts a bit, choose its opposite as $(\theta_b)_1$; if it predicts a nonbinary symbol, either bit causes a mistake. Fix the remaining coordinates of θ_b arbitrarily and reveal the resulting CoT trajectory. Independent choices across blocks leave one parameter vector $\theta \in B^d$ consistent with all revealed traces and force one mistake per block. Hence d mistakes are unavoidable. For a deterministic learner these deferred choices define one final θ and one transcript. Fixing that θ before play reproduces the same history and predictions, so the argument respects the fixed-target protocol.

Feedback and learner costs. Table 3 accounts for the explicit learners above. An endpoint learner stores an incidence vector for each S_b; a trajectory learner stores one address and a known/unknown flag per block. Storage counts these bits beyond the fixed prompt dictionary. The work bounds include input parsing and streamed feedback receipt, using words that hold token and input indices. With $\ell = \lceil \log_2 K \rceil$, fixed-width token coding uses ℓ bits per symbol; a binary endpoint also admits a one-bit encoding. A first prompt in an unknown block reveals one coordinate or all r address bits, while the target description has dr bits under either protocol (Proposition D.1). On prompt-only queries, coordinate masks also reduce endpoint state to $O(dr)$ bits.

Table 3: Resource costs and optimal mistake bounds for the explicit learners on arbitrary input words, $M \geq 2$. Feedback bits use fixed-width $\lceil \log_2 K \rceil$-bit token coding; a binary endpoint also admits a one-bit encoding.

Feedback	Mistakes	Feedback bits (fixed-width)	State bits	Work per round
Final	dr	ℓ	$O(d\|B\|)$	$O(\|w\| + \|B\|)$
Trajectory	d	$M\ell$	$O(d(r+1))$	$O(\|w\| + M)$

C.2 Fixed-Order Littlestone Tree

Bijectivity of R_M lets one fixed coordinate order realize every binary path.

Enumerate X_0 as $\xi_1, \ldots, \xi_D$. Form a complete binary tree of depth D, query ξ_t at every node of level t, and label every edge pair by $0, 1$. Every path is one binary labeling of X_0, and bijectivity of R_M supplies its realizing end-to-end map. This proves the Littlestone lower bound and yields the standard adversary: after the learner predicts, follow the opposite binary edge. A prediction in $\Sigma \setminus \{0, 1\}$ is also a mistake, and either binary edge may be chosen. After D rounds the path remains realizable.

C.3 Binary-Cube PAC Lower Bound

The confidence boundary case $D = 1$ is covered by the following standard form of the realizable lower bound.

Proposition C.1. *There is a universal $c > 0$ with the following property. Let $\mathcal{H} \subseteq \mathcal{Y}^{\mathcal{X}}$, suppose $\mathcal{X}$ contains distinct points $\xi_1, \ldots, \xi_D, \xi_*$, $D \geq 1$, and suppose $\mathcal{H}$ contains a designated family $\{h_\sigma : \sigma \in \{0,1\}^D\}$ such that*

$$h_\sigma(\xi_i) = \sigma_i \quad (i \in [D]), \qquad h_\sigma(\xi_*) = 0.$$

For $0 < \varepsilon \leq 1/8$ and $0 < \delta \leq 1/100$, every distribution-free learner, proper or improper, that is (ε, δ)-PAC in the realizable case has worst-case sample size at least

$$c\,\frac{D + \ln(1/\delta)}{\varepsilon}.$$

Proof. All target labels in the argument are binary. Map every nonbinary prediction of an improper multiclass learner to 0. This postprocessing cannot increase its error on the distributions below, so it suffices to prove the binary lower bound for the designated witness family.

The classical shattered-set lower bound gives $\Omega(D/\varepsilon)$ samples at any fixed confidence below $1/2$ (Ehrenfeucht et al., 1989; Hanneke, 2016). For $D = 1$, the same order follows directly from the two-hypothesis argument below with $\delta = 1/100$. Because a learner satisfying any $\delta \leq 1/100$ must also meet the fixed-confidence requirement at $1/100$, the dimension lower bound applies throughout the stated range.

For the confidence term, choose two witness hypotheses that differ at ξ_1 and agree at ξ_*, and put probability 2ε on ξ_1 and $1 - 2\varepsilon$ on ξ_*. For a sample of size n, conditional on observing no ξ_1, the labeled sample has the same distribution under both targets. Therefore for at least one target, even an improper randomized learner predicts the wrong label at ξ_1 with conditional probability at least $1/2$. On that event its population error is at least $2\varepsilon > \varepsilon$, so its failure probability is at least

$$\tfrac{1}{2}(1 - 2\varepsilon)^n.$$

To make this at most δ, one needs

$$n \geq \frac{\ln(1/(2\delta))}{-\ln(1-2\varepsilon)} \geq \frac{3}{8\varepsilon} \ln \frac{1}{2\delta},$$

where $-\ln(1-u) \leq u/(1-u)$ and $u = 2\varepsilon \leq 1/4$. Thus $n = \Omega(\ln(1/\delta)/\varepsilon)$. The maximum of this bound and the dimension bound is at least half their sum, proving the proposition after adjusting a universal constant.

C.4 Proof of the PAC Feedback Separation

Proof of Corollary 4.4. For end-to-end feedback, the common anchor required by Proposition C.1 is the empty prompt ϵ. Every generator first emits 0 there. Thereafter the contexts are $0, 00, \ldots$; they cannot be prompts because every prompt ends in 1, and they cannot be informative one-bit rays. Hence

$$f_\theta^{\mathsf{e2e},M}(\epsilon) = 0 \qquad \text{for every } \theta \text{ and } M \geq 1.$$

Proposition C.1 therefore applies to the D-point end-to-end cube, including when $D = 1$, and gives the lower bound.

For the matching end-to-end upper bound, realizability supplies a consistent hypothesis. A fixed hypothesis with error greater than ε matches n independent samples with probability at most $(1-\varepsilon)^n \leq e^{-\varepsilon n}$. A union bound over the 2^D end-to-end maps shows that

$$n \geq \frac{D \ln 2 + \ln(1/\delta)}{\varepsilon}$$

suffices, even if the learner is allowed to be improper.

For CoT feedback, fix the target parameter θ. The endpoint-one inputs associated with block b form

$$A_b(\theta) := \{x_{b,j} : (\theta_b)_j = 1\} \cup \bigcup_{j \in [r]:(\theta_b)_j = 1} \mathcal{R}_{b,j,\theta_b}.$$

The sets $A_b(\theta)$ are disjoint over b, and all other inputs associated with block b have endpoint 0. Initialize $\widehat{\theta}_b = 0^r$. Whenever a sample from block b has endpoint 1, its trace determines θ_b: a prompt trace begins with a_{θ_b}, while a positive routed-ray input contains the matching address. Set $\widehat{\theta}_b = \theta_b$ and return $f_{\widehat{\theta}}^{\mathsf{e2e},M}$ after processing the sample.

Let $p_b = P(A_b(\theta))$, and let U be the blocks from which no endpoint-one sample was observed. The returned predictor is exact on every learned block and predicts 0 throughout each block in U, so its risk is $\sum_{b \in U} p_b$. For a fixed $S \subseteq [d]$, write $p(S) = \sum_{b \in S} p_b$. If $p(S) > \varepsilon$, then

$$\Pr(S \subseteq U) = (1 - p(S))^n \leq e^{-n\varepsilon}.$$

If the risk exceeds ε, the realized unseen set U is one such subset. A union bound over at most 2^d subsets gives

$$\Pr\Big(P\big(f_{\widehat{\theta}}^{\mathsf{e2e},M}(x) \neq f_\theta^{\mathsf{e2e},M}(x)\big) > \varepsilon\Big) \leq 2^d e^{-n\varepsilon}.$$

Thus $(d \ln 2 + \ln(1/\delta))/\varepsilon$ samples suffice under CoT feedback.

For the matching lower bound, let $u = 0^r$ and $v = (1, 0, \ldots, 0)$, and restrict each block address to $\theta_b \in \{u, v\}$. At $\xi_b = x_{b,1}$, the two choices have endpoints 0 and 1, and their traces are

$$(a_u, 0, \ldots, 0) \qquad \text{and} \qquad (a_v, 1, \ldots, 1),$$

respectively. A labeled sample at ξ_b therefore reveals exactly the binary choice in block b, while samples at other $\xi_{b'}$ reveal no information about it. The empty prompt has the common trace 0^M. The binary-cube proof of Proposition C.1, now with $D = d$ and the two paired trace labels above, gives $\Omega((d + \ln(1/\delta))/\varepsilon)$ samples for every learner, including improper learners.

D A FINITE-STATE REALIZATION

The base class admits an explicit deterministic Moore realization over Σ. To characterize its state count, evaluation time, and description length, consider $(Q, q_{\mathrm{in}}, \Delta, \lambda)$, where Q is finite, $q_{\mathrm{in}} \in Q$, $\Delta : Q \times \Sigma \to Q$, and $\lambda : Q \to \Sigma$. Let $\Delta^* : Q \times \Sigma^* \to Q$ denote the word extension of Δ. The machine computes

$$f(w) = \lambda(\Delta^*(q_{\mathrm{in}}, w)).$$

Proposition D.1. *Let $L = 1 + \lceil \log_K(dr) \rceil$ be as in (7). Every $f_\theta \in \mathcal{F}$ is computed by a deterministic Moore machine with at most*

$$drL + 3 = dr\big(1 + \lceil \log_K(dr) \rceil\big) + 3 \tag{13}$$

states. The state set and prompt-trie skeleton are common to the entire class. Given a transition table, evaluation of $f_\theta(w)$ takes exactly one transition per symbol of w, hence $O(|w|)$ time.

With the shared alphabet and indexed prompt dictionary fixed, each generator has a dr-bit identifier, and no shorter fixed-length identifier can distinguish all class members.

Proof. Let P contain ϵ and every nonempty prefix of a designated prompt. Since there are dr prompts, each of length L,

$$|P| \le 1 + drL.$$

Add two states q_{one} and q_{dead}. The initial state is ϵ. At a proper prompt prefix $v \in P$, reading $c \in \Sigma$ moves to vc if $vc \in P$, and otherwise moves to q_{dead}. At a complete prompt $x_{b,j}$, reading c moves to q_{one} precisely when

$$c = a_{\theta_b} \quad \text{and} \quad (\theta_b)_j = 1;$$

all other symbols move to q_{dead}. The state q_{one} loops on 1 and moves to q_{dead} on every other symbol; q_{dead} loops on every symbol.

Set the Moore output of $x_{b,j}$ to a_{θ_b}, the output of q_{one} to 1, and every other state output to 0. An input ending at a prompt is therefore handled by the first clause of (3). A matching routed ray with selected bit 1 enters q_{one} and remains there exactly while further 1's are read. Every matching ray whose selected bit is 0, every nonmatching ray, and every word outside the designated sets has output 0, exactly as in the remaining clauses of (3). Thus the machine computes f_θ, and its number of states is at most $|P| + 2 \le drL + 3$.

The concatenation $\theta_1 \cdots \theta_d$ is an injective dr-bit identifier and determines all member-specific prompt outputs and transitions above. Conversely, if fixed-length identifiers had length $p < dr$, they could name at most $2^p < 2^{dr} = |\mathcal{F}|$ distinct generators, contradicting injectivity.

Transition-table accounting. Let $Q_0 = drL + 3$. With state and alphabet symbols represented by integer indices, a dense representation of the transition and output tables uses at most

$$KQ_0\lceil \log_2 Q_0 \rceil + Q_0 \lceil \log_2 K \rceil$$

bits. A succinct representation instead stores the common indexed prompt trie once, then uses the optimal dr-bit identifier to instantiate the member-specific terminal outputs and matching-route transitions.

E EXPERIMENTAL DETAILS

Tokens are represented as integers, and prompts use fixed-width base-K encoding. Exact searches use $d \in \{1, 2, 3\}$, $4 \le K \le 8$, and $M \in \{1, 2\}$. The factorized alphabet and horizon sweep fixes $d = 3$ and uses $4 \le K \le 130$ and $M \in \{1, 2, 4, 8, 16\}$.

E.1 DIMENSION SEARCH AND ONLINE LEARNING

For a finite hypothesis class, the Littlestone search applies the standard fiber recursion with memoized version-space bitmasks. The Natarajan search keeps one nonempty version cell for every binary assignment to the selected label pairs; a new instance is admissible only if it splits every existing cell.

For a finite class, the product of its Moore machines provides a finite representation of input words. Words reaching the same product state have identical labels under every hypothesis, and breadth-first traversal enumerates the reachable states.

At each query, SOA chooses a feedback label whose consistent version-space fiber has maximum exact Littlestone dimension and predicts its endpoint; majority predicts the endpoint supported by the most remaining hypotheses. We evaluate both possible endpoint tie breaks. The trace-blind ablation generates a full trajectory but conditions the learner only on its endpoint. For the alphabet sweep, the factorized learner tracks unresolved address bits or unresolved blocks directly, avoiding enumeration of $\mathcal{F}$ while retaining the same version-space updates on prompt queries.

For endpoint and trace-blind feedback, the online adversary sets each unresolved coordinate opposite to the learner's current prediction. Under trajectory feedback, it instead chooses one such address when a block is first queried; the revealed trajectory then fixes the rest of that block. After the addresses are fixed, the learner is run again on the same query sequence against that target.

Table 4 reports the fixed-target dynamics shown in Figure 3(a). Endpoint-only updates halve the version space at every coordinate. Under trajectory feedback, the first informative update to an unresolved block fixes its whole address, while the second coordinate adds neither information nor another mistake.

Table 4: Version-space dynamics for the fixed all-one target at $d = 3, K = 8, M = 2$, using SOA with the 0-first tie rule. $|V_t|$ is measured after update t; Final and trace-blind coincide, and Majority follows the same sequence on this symmetric run.

Query		Version-space size		Cumulative mistakes	
t	(b, j)	Final / trace-blind	Trajectory	Final / trace-blind	Trajectory
0	–	64	64	0	0
1	$(1, 1)$	32	16	1	1
2	$(1, 2)$	16	16	2	1
3	$(2, 1)$	8	4	3	2
4	$(2, 2)$	4	4	4	2
5	$(3, 1)$	2	1	5	3
6	$(3, 2)$	1	1	6	3

E.2 PAC Learning Curves

The batch sweep fixes $M = 2$, with the empty prompt $x_\circ = \epsilon$ as a common zero-output anchor. Write $e_1 = (1, 0, \ldots, 0) \in B$. The protocol-specific dimension distributions are

$$P^{\mathsf{e2e}}_{\dim}(x_{b,j}) = \frac{4\varepsilon}{dr}, \qquad P^{\mathsf{CoT}}_{\dim}(x_{b,1}) = \frac{4\varepsilon}{d}, \qquad P^{\mathsf{e2e}}_{\dim}(x_\circ) = P^{\mathsf{CoT}}_{\dim}(x_\circ) = 1 - 4\varepsilon.$$

Their targets have $\theta_b = 1^r$ and $\theta_b = e_1$, respectively, for every block. For both protocols, the confidence distribution assigns 2ε to $x_{1,1}$ and $1 - 2\varepsilon$ to $x_\circ$, with $\theta_1 = e_1$ and all other addresses zero. All unlisted points have zero mass. The ERM selects the consistent parameter minimizing

$$I(\theta) = \sum_{b=1}^{d} \sum_{j=1}^{r} (\theta_b)_j \, 2^{(b-1)r+j-1}.$$

Consistency uses the observed feedback; unconstrained coordinates are set to zero.

For the dimension distribution, set $u = dr$ for final feedback and $u = d$ for trajectory feedback, and let $a = 4\varepsilon/u$. If $p_n(s)$ is the probability that s distinct informative units have appeared in n samples, then $p_0(0) = 1$, $p_0(s) = 0$ for $s > 0$, and

$$p_{n+1}(s) = [1 - (u - s)a]p_n(s) + (u - s + 1)a\, p_n(s - 1), \qquad 0 \le s \le u,$$

with out-of-range states zero. The endpoint risk is $a(u - s)$, giving

$$F_{\dim}(n) = \sum_{s:\, a(u-s)>\varepsilon} p_n(s), \qquad F_{\mathrm{conf}}(n) = (1 - 2\varepsilon)^n.$$

Each solid-curve threshold is the smallest n for which $\max\{F_{\dim}(n), F_{\mathrm{conf}}(n)\} \le \delta$.

Figure 4 varies the address dimension, target error, and confidence. The final-feedback curve follows the number of readable address bits, whereas the trajectory curve follows the number of blocks. Both curves share the confidence contribution. Direct enumeration on small shared-address classes checks that fitting the ERM from actual endpoint or trajectory labels gives the same unseen-mass risk as the dynamic program.

The settings in Figure 4 are as follows. **(a)** At $d = 3$, $\varepsilon = 1/16$, and $\delta = 10^{-2}$, increasing the alphabet raises the final-feedback dimension term while the trajectory threshold stays fixed. **(b)** At $d = 3, K = 10, \delta = 10^{-2}$, both thresholds scale inversely with target error. **(c)** At $d = 3, K = 10, \varepsilon = 1/16$, tightening confidence adds a common logarithmic cost.

For a paired comparison, fix $\theta_b = 1^r$ for every block and use $P(x_{b,j}) = 4\varepsilon/(dr)$ and $P(x_\circ) = 1 - 4\varepsilon$ under both feedback protocols. With the indexed ERM above, let U_n and V_n count the unseen coordinates and unseen blocks after the same n samples. The endpoint risks are

$$R_n^{\mathsf{e2e}} = \frac{4\varepsilon U_n}{dr}, \qquad R_n^{\mathsf{CoT}} = \frac{4\varepsilon V_n}{d}.$$

An unseen block contains r unseen coordinates, so $R_n^{\mathsf{CoT}} \le R_n^{\mathsf{e2e}}$ on every sample sequence. Each block has total mass $4\varepsilon/d$; the occupancy recurrence therefore applies with $u = dr$ and $u = d$, respectively, giving Table 5.

Table 5: Paired PAC thresholds under the common distribution and all-one target, with $d = 3, M = 2, \varepsilon = 1/16, \delta = 10^{-2}$. Entries are the smallest n satisfying $\Pr(R_n > \varepsilon) \le \delta$ for the fixed ERM.

K	r	Final	Trajectory
6	2	84	66
18	4	120	66
66	6	180	66

E.3 Moore Machine Evaluation

The Moore-machine measurements include $d \in \{1, 3, 8\}$ and every alphabet size from four through sixteen. Classes with at most 64 members are exhaustive; larger classes use 16 sampled hypotheses. Structured ray contexts are combined with random words, including configurations whose prompts have length three. We remove unreachable states and apply Moore partition refinement before reporting the minimized count. The minimized machines show how much pruning reduces the explicit construction, while the direct evaluator supplies the independent output comparison.

On the tested words, the direct generator and Moore-machine outputs agree. The constructed state counts satisfy Proposition D.1, and transition counts equal input lengths.

Figure 5 aggregates the full grid, separating the explicit construction from the smaller machines found after reachability pruning and partition refinement.

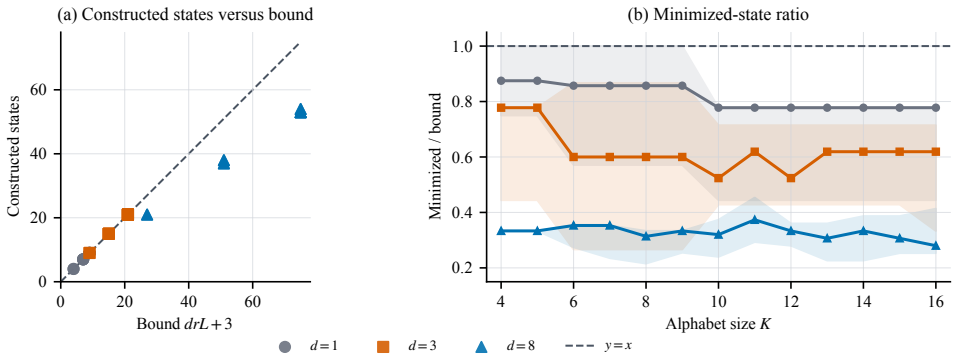


Figure 5: Finite-state size comparison. **(a)** Points represent (d, K) configurations; constructed state counts lie on or below the proposition's $y = x$ bound. **(b)** Curves show the median ratio of minimized states to the bound, and shaded bands show the full range over the hypotheses evaluated at each configuration. Families with at most 64 members are exhaustive; the others use 16 sampled hypotheses.

Table 6 resolves the same comparison at the first three routing thresholds, which are the thresholds inside the audited alphabet range. The explicit built count equals the upper-bound value in several exhaustive rows, while shared prompt prefixes and machine minimization produce instance-dependent reductions elsewhere.

Table 6: Moore-machine counts at the first three routing thresholds. State ranges are over the tested hypotheses. Scope E exhausts the indicated hypothesis family; scope S evaluates 16 sampled hypotheses. Percentage ranges are minimized states divided by $drL + 3$, rounded to the nearest percentage point.

Configuration				**Hypotheses**		**State count**				
d	K	r	L	$\#\theta$	**Scope**	**Built**	**Reachable**	**Minimized**	**Bound**	**Min./bound (%)**
$d = 1$										
1	4	1	1	2	E	4	3–4	3–4	4	75–100
1	6	2	2	4	E	7	6–7	4–7	7	57–100
1	10	3	2	8	E	9	8–9	4–7	9	44–78
$d = 3$										
3	4	1	2	8	E	9	8–9	4–7	9	44–78
3	6	2	2	64	E	15	14–15	4–13	15	27–87
3	10	3	2	16	S	21	21–21	9–15	21	43–71
$d = 8$										
8	4	1	3	16	S	21	21–21	9–9	27	33–33
8	6	2	3	16	S	38	38–38	14–18	51	27–35
8	10	3	3	16	S	54	54–54	18–28	75	24–37

F Constants and Thresholds

The two-step alphabet bound and the joint envelope use explicit constants and thresholds. This section verifies those constants over the full alphabet range and the large-horizon regime stated in the corresponding corollaries.

The two-step statement used in the main text is

$$\mathfrak{E}_2(d, K) = \Theta(d \log K), \qquad \frac{d}{3} \log_2 K \le \mathfrak{E}_2(d, K) < 6d \log_2 K. \tag{14}$$

Lower-bound constant. For $4 \le K < 8$, direct evaluation gives $\lfloor \log_2(K-2) \rfloor \ge 1 > \frac{1}{3}\log_2 K$. For $K \ge 8$,

$$\lfloor \log_2(K-2) \rfloor \ge \log_2 K - 2 \ge \tfrac{1}{3}\log_2 K.$$

Thus the lower constant in (14) covers the entire stated alphabet range, including $K = 4$.

Upper constant at two steps. For $K \ge 4$, the ratio $K^3/(K-1)$ is increasing and

$$\frac{4^3}{4-1} = \frac{64}{3} > 4e.$$

Hence $4e(K-1) < K^3$, so

$$2\log_2\big(4e(K-1)\big) < 6\log_2 K.$$

This is the upper constant in (14).

Joint-envelope thresholds.

Corollary F.1 (Joint Alphabet–Horizon Envelope)**.** *There exist a universal integer $d_\star$ and $M_\star : \mathbb{N} \to \mathbb{N}$ such that $\mathfrak{E}_M(d, K) = \Theta(d \log(KM))$ for all integers $d \ge d_\star$, $K \ge 4$, and $M \ge M_\star(d)$.*

Proof. The binary construction of Doron-Arad et al. (2026, Theorem 8.2, via Lemmas 8.10 and 8.13) supplies universal $c, C > 0$, an integer d_0, and a threshold function $M_0 : \mathbb{N} \to \mathbb{N}$ such that, for every $d' \ge d_0$, there is a binary base class $\mathcal{G}_{d'}$ with

$$\mathrm{Ldim}(\mathcal{G}_{d'}) \le Cd', \qquad \mathrm{Ldim}(\mathcal{G}_{d'}^{\mathsf{e2e},M}) \ge cd' \log_2 M \quad \text{for every } M \ge M_0(d').$$

The second inequality follows because the induced class VC-shatters a fixed set of at least $cd' \log_2 M$ inputs. Querying that set in a fixed order gives the same end-to-end Littlestone lower bound. Increase C if necessary so that $C \ge 1$, take $d_\star = \lceil 2C(d_0+1) \rceil$, define $M_\star(d) = 2$ arbitrarily for $d < d_\star$, and, for $d \ge d_\star$, set

$$d' = \lfloor d/C \rfloor, \qquad M_\star(d) = \max\{2, M_0(d')\}.$$

Then $d' \ge d_0$, $Cd' \le d$, and the end-to-end lower bound is at least $(c/2C)d \log M$.

Embed this class in a K-token alphabet by making every generator output 0 on words containing an added token. The extensions agree on all such words, so the base-class Littlestone dimension is unchanged; trajectories from binary prompts remain binary, preserving the end-to-end lower bound. Theorem 4.1 also gives $\mathfrak{E}_M(d, K) \ge \frac{1}{3} d \log_2 K$ for every $M \ge 2$. Both the shared-address class and the embedded binary class are admissible in the envelope. With $c_0 := \min\{1/3, c/(2C)\}$,

$$\mathfrak{E}_M(d, K) \ge c_0 d \max\{\log_2 K, \log_2 M\} \ge \frac{c_0}{2} d \log_2(KM).$$

Theorem 3.3 gives the matching upper bound.

G A Six-Symbol Example

Take $d = 1$ and $K = 6$, so $r = 2$, $B = \{00, 01, 10, 11\}$, and no filler symbols are needed: $\Sigma = \{0, 1, a_{00}, a_{01}, a_{10}, a_{11}\}$. Formula (7) gives $L = 2$; choose the two distinct prompts $p_j := x_{1,j}$, both ending in 1. For $s \in B$, write $f_s := f_{(s)}$. Then $f_s(p_j) = a_s$, while $f_s^{\mathsf{e2e},M}(p_j) = s_j$ for every $M \ge 2$.

Table 7: Six-symbol readout at $d = 1, K = 6$. The routing token a_s identifies the address, and p_1, p_2 read its two bits for every $M \ge 2$, realizing all four binary labelings.

Hidden routing		**Final token** $(M \ge 2)$	
Address s	**Address token** a_s	p_1	p_2
00	a_{00}	0	0
01	a_{01}	0	1
10	a_{10}	1	0
11	a_{11}	1	1

Thus both end-to-end dimensions are two, while the base and CoT dimensions are one.