MINT: CERTIFYING QUEUE INSTABILITY FROM LIMITED COMMANDS AND HIDDEN MODES

Anonymous authors
Paper under double-blind review

ABSTRACT

Can observation and command interfaces each pass a capacity test while together making stabilization impossible? We introduce *MINT* (Menu–Innovation No-go Test), a model-based certificate for controllers that send one of K labels to a stateless executor with fixed action templates. MINT first averages each action's service over the unseen mode change, then finds the best service possible with K fixed action templates, even granting the controller knowledge of the previous mode. If arrivals exceed this joint ceiling, every compliant controller using past departure feedback has growing expected backlog, regardless of its memory or training method. Separate command-only and prediction-only ceilings can both miss this obstruction. For every $m \geq 3$, we give an m-mode queue family and an open load interval where MINT detects it. Matching stable constructions establish exact minimum budgets of m commands with departure history and $m-1$ with current-mode observation; two noncyclic instances show the same separation. In a four-mode task, reward-only search recovers finite controllers that are certified stable after training at a feasible budget.

1 INTRODUCTION

Observation and command capacity must be evaluated together: two separate service bounds can both exceed demand while their joint constraint makes every compliant controller unstable. We introduce *MINT* (Menu–Innovation No-go Test) to detect this obstruction, and construct queues whose exact minimum command budgets differ under departure feedback and current-mode observation. Matching stable controllers turn the class-wide exclusion into a precise interface-design threshold.

We use a queue in which jobs arrive while a hidden server mode (or *phase*) determines which action is likely to serve them. The controller learns whether a job departed after each attempt, but the mode can change before its next decision. Even complete feedback history cannot reveal that fresh change. A persistent gap between arrivals and the best possible service makes unfinished work accumulate.

At execution, the history processor sends one of K labels, or command IDs, per service attempt. A stateless executor sees only the current label and draws a physical action from its installed probability distribution, called an *action template*. The K templates together form a *menu*; the executor is the decoder that maps a label to one template. For example, a recurrent dispatcher may compress history into a command ID sent through a finite API; the executor then runs one installed template. Each template may use every physical action. Longer history can change the chosen label, whereas a larger K makes more templates selectable. Sending a full state estimate for the executor to interpret, or letting it remember prior messages, changes this interface. Thus K counts installed templates, not physical actions, memory states, or encoded bits. This label-only execution contract is an explicit architectural assumption. Figure 1 shows the contract.

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.

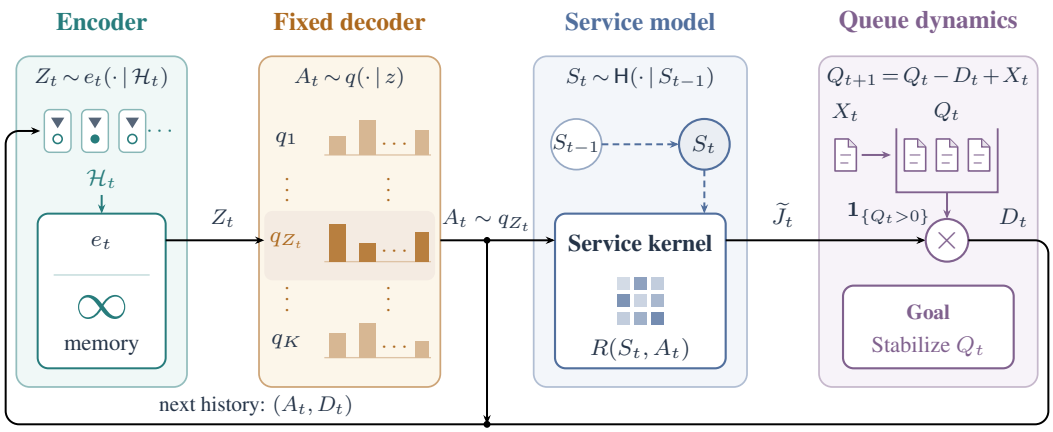

Figure 1: **The label-only control contract tested by MINT.** An unrestricted encoder selects $Z_t \in [K]$ from past actions and departures. The fixed decoder samples A_t from the selected row using fresh private randomness; each row may support all actions. Potential service produces departures only from nonempty queues. Current state and queue length are unobserved; distribution bars are schematic. For the family and loads in Theorem 4, MINT excludes every compliant $(m-1)$-label departure-history controller, regardless of memory. Increasing K to m, or revealing S_t while retaining $K = m-1$, admits a positive-recurrent controller.

Service-success feedback in partially observed servers is studied by Asanjarani & Nazarathy (2020). Finite-action policy approximation and data-rate-limited control motivate studying constrained control interfaces (Saldi et al., 2015; Tatikonda & Mitter, 2004; Nair & Evans, 2004).

MINT's key step is to evaluate the command menu *after* accounting for the unseen mode change. We compare it with two simpler necessary service bounds. The interface-only oracle sees the current phase but retains K labels. The prediction-only oracle knows the previous phase and may select any action. The joint oracle knows the previous phase but retains K labels and cannot see the fresh change. Both separate bounds can exceed demand while MINT's ceiling lies below it. Under the stated assumptions, MINT then excludes the entire interface class. Our matched-cycle family makes the resulting command threshold exact through matching stable controllers.

Contributions.

1. **A joint obstruction missed by separate tests.** Given a fresh hidden-mode transition, service matrix, and K-command interface, MINT computes a joint ceiling on service. When arrivals exceed it, every compliant departure-history controller has growing expected backlog, regardless of memory or training method (Theorem 2); compliant homogeneous realizations have no invariant law.

2. **An exact command threshold.** For every $m \geq 3$, an m-mode cycle has an open load interval where MINT excludes a controller class that neither separate bound can exclude. Matching stable controllers establish exact minimum budgets of m commands with departure feedback and $m-1$ with current-mode observation (Theorem 4). The constructions include the empty-queue boundary; a higher-load result also gives exponential queue tails. For general finite models with current-phase observation and fresh interior primitives, we characterize the minimum stable menu; deciding whether a budget suffices is NP-complete (Appendix A).

3. **Certified learning at a feasible budget.** On the four-phase task, reward-only search recovers finite controllers whose stability is certified after training. Neural runs show how training choices affect finite-run performance within an adequate interface.

2 MODEL AND THE MINT CERTIFICATE

We first specify the execution contract, then derive its service ceiling and queue consequences.

Queue and observations. Let S_t take values in a finite set $\mathcal{S}$ and evolve as an autonomous Markov chain with transition matrix H. Write $\nu_t = \mathcal{L}(S_t)$ for its probability distribution, so $\nu_t = \nu_{t-1}\mathsf{H}$ for $t \geq 1$. Actions belong to a finite set $\mathcal{A}$. Conditional on $(S_t, A_t) = (s, a)$, potential service $\widetilde{J}_t \in \{0, 1\}$ has mean $R(s, a) \in [r_{\min}, r_{\max}] \subseteq [0, 1]$. Independent Bernoulli arrivals X_t have mean λ, and

$$D_t = \mathbf{1}\{Q_t > 0\}\widetilde{J}_t, \qquad Q_{t+1} = Q_t - D_t + X_t. \tag{1}$$

Potential service $\widetilde{J}_t$ is a possible completion; actual departure D_t is zero when the queue is empty, even if that service would have succeeded. The public history $\mathcal{H}_t$ contains past actual departures and the controller's past actions, but neither S_t nor Q_t. We call a homogeneous augmented controller stable when its reachable queue–controller–phase chain is positive recurrent. Our explicit constructions have finite controller state. For arbitrary history policies we state finite-time expected-backlog bounds; invariant-law exclusions concern compliant homogeneous augmented Markov realizations.

2.1 THE ACTION-TEMPLATE INTERFACE

No-bypass K-label controller. Write $[K] = \{1, \ldots, K\}$. At decision time an arbitrary randomized, possibly nonstationary history encoder emits $Z_t \in [K]$; fresh action randomness then passes through a kernel that receives only Z_t. One label specifies one service attempt. Sending a state estimate for the executor to interpret, or letting it combine messages across slots, would give it a different input or memory and is outside this contract. For every feasible public history h,

$$\Pr(Z_t = z, A_t = a \mid \mathcal{H}_t = h) = e_t(z \mid h)q_t(a \mid z). \tag{2}$$

We additionally require the same factorization after conditioning on S_t: $\Pr(Z_t = z, A_t = a \mid \mathcal{H}_t = h, S_t = s) = e_t(z \mid h)q_t(a \mid z)$. This additional condition rules out information bypasses; it does not follow from (2) alone. The encoder cannot read the current hidden state. Beyond the label, the decoder receives no hidden state, queue length, full history, or shared random signal that could carry such information. History may nevertheless correlate S_t with Z_t. The finite-time impossibility bounds even allow prearranged changes to the decoder rows over time. Invariant-law claims require one fixed action law $q(a \mid z)$ even after conditioning on any augmented encoder or clock state: an action-visible clock or decoder memory is an additional resource.

For a fixed deployed decoder,

$$\pi(a \mid h) = \sum_{z=1}^{K} e(z \mid h)q(a \mid z).$$

History chooses a point in the convex hull of K action-distribution templates, each possibly supported on every physical action. More encoder memory changes the history-to-mixture map; more labels enlarge the available library. History-dependent experts and nonlinear mixtures of logits need separate analysis (Appendix D).

2.2 MINT: TESTING THE MENU AND MODE CHANGE TOGETHER

Each service bound below is a necessary test: demand above it rules out the stated controller class; a bound above demand alone does not supply an implementable stable controller. For a hidden-state law ν, define the full-information menu capacity

$$\Gamma_K(\nu, R) = \max_{a_1, \ldots, a_K \in \mathcal{A}} \sum_{s \in \mathcal{S}} \nu(s) \max_{j \leq K} R(s, a_j). \tag{3}$$

It grants the encoder current-phase knowledge while retaining a label-only decoder. A deterministic action menu maximizes this linear upper-bound calculation; the controller class itself still permits randomized decoder rows over every action. We call the resulting necessary test the *interface-only bound*. Clearing it does not establish stability under departure feedback. With a finite irreducible phase

chain, fresh independent draws, and interior service probabilities, current-phase observation makes this threshold exact: a stable K-label controller exists precisely when $\lambda < \Gamma_K(\pi, R)$ (Proposition 5). Appendix A also gives the menu number, bit interpretation, and computational complexity.

For the interface-only comparison, the encoder may additionally see S_t: its conditional law may be $e_t(z \mid h, s)$, but the decoder still uses only $q_t(a \mid z)$. We use $\mathrm{TV}(\nu, \pi) = \sup_B |\nu(B) - \pi(B)| = \frac{1}{2}\|\nu - \pi\|_1$.

Theorem 1 (Interface-only action-menu bound)**.** *Every K-label controller whose state-conditioned action law factors as $\Pr(Z_t = z, A_t = a \mid \mathcal{H}_t = h, S_t = s) = e_t(z \mid h, s) q_t(a \mid z)$ satisfies*

$$\mathbb{E}\widetilde{J}_t \le \Gamma_K(\nu_t, R), \qquad \nu_t = \mathcal{L}(S_t). \tag{4}$$

If $\mathbb{E}Q_0 < \infty$, then for every $T \ge 0$,

$$\mathbb{E}Q_T \ge \mathbb{E}Q_0 + \sum_{t=0}^{T-1} \big(\lambda - \Gamma_K(\nu_t, R)\big). \tag{5}$$

For every reference law π,

$$\Gamma_K(\nu_t, R) \le \Gamma_K(\pi, R) + (r_{\max} - r_{\min})\,\mathrm{TV}(\nu_t, \pi). \tag{6}$$

If the hidden chain has unique invariant law π and $\lambda > \Gamma_K(\pi, R)$, no compliant homogeneous full augmented realization has an invariant probability law.

Proof sketch. Let $\xi_t(s, z) = \Pr(S_t = s, Z_t = z)$. Choosing one maximizing action per non-null label gives

$$\mathbb{E}\widetilde{J}_t \le \sum_z \max_a \sum_s \xi_t(s, z) R(s, a) \tag{7}$$

$$\le \sum_s \nu_t(s) \max_{z \le K} R(s, a_z) \le \Gamma_K(\nu_t, R). \tag{8}$$

For each label, the bound assigns the action with the highest service averaged over states producing that label. This inner sum has the finite-message posterior-gain form of Alvim et al. (2012); optimizing over K labels yields our menu expression. It preserves arbitrary state–label correlation. Queue telescoping and bounded stationary flow balance yield the remaining claims (Appendix A). Thus the interface-only bound also covers current-phase encoders; the joint theorem below uses the stricter no-bypass contract.

The *prediction-only bound* grants the encoder the previous hidden state and enough labels to select any physical action. It still cannot observe an independent current innovation. For $t \ge 1$, set

$$(\mathsf{H}R)(s, a) = \sum_{s'} \mathsf{H}(s, s') R(s', a), \qquad B_{\mathrm{pred},t} = \sum_s \nu_{t-1}(s) \max_a (\mathsf{H}R)(s, a).$$

This is an oracle upper bound, not a rate attained by departure feedback. When only K labels are available, MINT combines the two limits: it first computes the effective service matrix $\mathsf{H}R$, averaging over the mode change that the controller cannot yet see, and then optimizes the K-slot action menu. Its *joint ceiling* is $\Psi_K(\nu, \mathsf{H}, R) = \Gamma_K(\nu, \mathsf{H}R)$. For an invariant hidden-mode law π, MINT issues a no-go certificate when $\lambda > \Psi_K(\pi, \mathsf{H}, R)$; otherwise the test is inconclusive. This is a finite, model-based calculation, not a trained policy or a general-purpose stability oracle. The theorem makes its verdict precise.

Theorem 2 (MINT joint menu–innovation bound)**.** *For every $t \ge 1$, suppose the no-bypass controller additionally obeys, at all feasible conditioning values,*

$$\Pr(S_t = s' \mid S_{t-1} = s, Z_t = z) = \mathsf{H}(s, s'), \tag{9}$$

$$\Pr(A_t = a \mid S_{t-1}, S_t, Z_t = z) = q_t(a \mid z). \tag{10}$$

Then

$$\mathbb{E}\widetilde{J}_t \le \Psi_K(\nu_{t-1}, \mathsf{H}, R) \le \min\{\Gamma_K(\nu_t, R), B_{\mathrm{pred},t}\}. \tag{11}$$

If $\mathbb{E}Q_0 < \infty$, for every $T \geq 1$,

$$\mathbb{E}Q_T \geq \mathbb{E}Q_0 + \lambda - \Gamma_K(\nu_0, R) + \sum_{t=1}^{T-1} \big[\lambda - \Psi_K(\nu_{t-1}, \mathsf{H}, R)\big]. \tag{12}$$

If the hidden chain has unique invariant law π and $\lambda > \Psi_K(\pi, \mathsf{H}, R)$, no compliant homogeneous full augmented realization satisfying these conditional laws has an invariant probability law.

Proof sketch. Conditioning on (S_{t-1}, Z_t) averages the unobserved transition into $\mathsf{H}R$; the same menu argument proves the first inequality. Optimizing without a label limit gives $B_{\mathrm{pred},t}$, while averaging a menu after the transition gives $\Gamma_K(\nu_t, R)$. Slot zero uses only Theorem 1; telescoping gives (12). Appendix E proves the TV and invariant-law steps. Autonomy alone does not imply (9); the product primitives of our construction below ensure both conditional laws. Granting the encoder current-phase observation generally breaks (9). The same calculation handles older information: under the corresponding delayed-information laws, each fixed-budget ceiling is nonincreasing with information age, and mixing sends all budgets to the same blind-service limit (Corollary 21).

3 Where MINT is tight: exact budgets and interventions

3.1 A matched-cycle family

For the central separation, take $m \geq 3$ and $\mathcal{S} = \mathcal{A} = \mathbb{Z}_m$. The hidden phase evolves as

$$S_{t+1} = S_t + H_t \pmod m, \qquad \Pr(H_t = 0) = p, \quad \Pr(H_t = 1) = 1 - p, \tag{13}$$

and potential service has mean

$$R(s, a) = b + (g - b)\mathbf{1}\{s = a\}, \qquad 0 < b < g < 1. \tag{14}$$

Given the full predecision past and (Q_t, S_t, A_t), the potential service, arrival, and hidden increment use independent fresh Bernoulli draws with parameters $R(S_t, A_t)$, λ, and $1 - p$. The event order is action, potential service, actual departure, arrival, then hidden transition. The controller observes only $Y_t = D_t$. Thus the label for slot t uses feedback through slot $t - 1$; the independent transition from S_{t-1} to S_t occurs afterward and cannot enter that label or the executor's fresh action draw. This product primitive law makes the innovation conditions in Theorem 2 valid for departure-history controllers; the general interface-only bound needs only its conditional service means.

Choose the physical model parameters

$$b, \kappa, \zeta \in [1/8, 1/4], \qquad p = 1 - \kappa/m, \qquad g = 1 - \zeta/(m-1)^2. \tag{15}$$

The arrival rate is selected below according to the stability question. The hidden chain has invariant law $u_m(s) = 1/m$. With $k = \min(K, m)$, the interface-only and joint bounds have closed forms

$$\begin{aligned} c_K &:= \Gamma_K(u_m, R) = b + (g - b)k/m, \\ \Psi_K &:= \Psi_K(u_m, \mathsf{H}_p, R) = b + \frac{g - b}{m}\big[pk + (1 - p)\min(k, m - k)\big]. \end{aligned} \tag{16}$$

The prediction-only bound is $B_{\mathrm{pred}} = \Psi_m = b + (g - b)p$. These are bounds on controller classes; only a specified construction can establish attainability. Appendix E proves the MINT closed form, and Appendix B records transient corrections for nonuniform hidden initialization.

3.2 A stabilizing departure-feedback tracker

For $K = m$, index the m labels by $\mathbb{Z}_m$ through a bijection with $[m]$. Starting from $Z_0 = 0$, the label-only decoder and the departure-feedback update are

$$A_t = Z_t, \qquad Z_{t+1} = Z_t + \mathbf{1}\{Y_t = 0\} \pmod m. \tag{17}$$

Here the stored phase guess and emitted label coincide; controller memory and labels are distinct resources in general. A departure favors retaining a matching action; without a departure the tracker searches the next phase. At an empty queue, however, non-departure does not reveal a failed potential service. The closed-loop proof treats that boundary separately.

Proposition 3 (Tracker service and stability). *For the physical family (15), the specified m-label tracker has saturated potential-service rate (its long-run potential service if work is always waiting) $\mu_{\mathrm{tr}} > c_{m-1}$, independently of arrivals. For every $0 < \lambda < \mu_{\mathrm{tr}}$, its queue–phase–tracker chain is positive recurrent and has exponential stationary queue tails.*

Proof sketch. In the saturated finite difference-phase chain, matches persist while mismatches return toward a match; its invariant match mass π_0 gives $\mu_{\mathrm{tr}} = b + (g-b)\pi_0 > c_{m-1}$. To turn this saturated calculation into a queue claim, account for phase-dependent drift and feedback at an empty queue. For phase kernel P and service vector r, solve $h - Ph = \mu_{\mathrm{tr}}\mathbf{1} - r$. At positive queue levels, $V(q,i) = q + h_i - \min_j h_j$ has drift $\lambda - \mu_{\mathrm{tr}} < 0$. Strictly interior primitive probabilities connect the finite empty-queue boundary, giving positive recurrence. Bounded increments give an exponential transform for each fixed strict-slack load. The excursion, drift, and boundary arguments are in Appendices D and F.

3.3 Exact thresholds under two observation regimes

Theorem 4 (Exact effective interface thresholds). *For every $m \geq 3$ and physical parameter choice in (15), let*

$$\mathcal{I}_m = (\Psi_{m-1}, c_{m-1}), \qquad |\mathcal{I}_m| = (g-b)\kappa(m-2)/m^2 = \Theta(1/m).$$

For every $\lambda \in \mathcal{I}_m$, the following hold in the same environment.

1. *At $K = m-1$, both separate interface-only and prediction-only bounds exceed λ, but MINT excludes every compliant departure-history controller with $K \leq m-1$, regardless of its upstream memory or computation.*
2. *The specified m-label tracker is positive recurrent under unchanged departure feedback, so the minimum stable label count under that observation contract is m.*
3. *Granting current-phase information to the encoder makes an $(m-1)$-label menu controller positive recurrent. The interface-only bound excludes every controller with $K \leq m-2$ even under this observation, so the minimum label count is $m-1$.*

Let $K^{\min}_{\mathrm{past}}$ and $K^{\min}_{\mathrm{phase}}$ denote the smallest label budgets admitting a positive-recurrent time-homogeneous controller under the two respective observation contracts. Then

$$K^{\min}_{\mathrm{past}} = m, \qquad K^{\min}_{\mathrm{phase}} = m-1.$$

For both exclusions, finite-mean initial queues have at least linear expected-backlog growth up to a finite initialization correction; compliant homogeneous augmented realizations have no invariant probability law.

Extra labels can be left unused, so the feasible budgets in this family and load interval are exactly $K \geq m$ under departure feedback and $K \geq m-1$ with current-phase observation.

Proof sketch. Structurally, the MINT ceiling falls below demand while both single-resource tests remain inconclusive; the extra-label tracker and current-phase menu then attain rates above demand. The closed forms give $c_{m-2} < \Psi_{m-1} < \lambda < c_{m-1} < \Psi_m$. Theorem 2 excludes $K \leq m-1$ under departure history, and Proposition 3 supplies the unchanged-feedback construction. For current-phase observation, a deterministic $(m-1)$-row menu matches all but one phase and attains c_{m-1}. Its finite-phase drift correction $h_{\mathrm{obs}}(s) = s(g-b)/\kappa$ gives positive-level drift $\lambda - c_{m-1} < 0$; the finite communicating boundary closes the recurrence proof. Theorem 1 excludes $K \leq m-2$. Appendix F gives the full proof, including transient and invariant-law claims.

Four-phase example. With three templates for four matched actions, one phase lacks its matching action. Even an oracle told the previous phase cannot see whether the mode stayed or jumped before choosing a label. For Figure 2, the menu $\{0,1,2\}$ therefore has predicted match probabilities $p, p, p, 1-p$ across the four previous phases: after the omitted previous phase, a match requires a jump to a covered action. This joint limit falls below demand even though both separate bounds lie above it. The figure compares those three necessary tests with two attained saturated rates. Exact endpoints and queue certificates are in Appendix F. The same exact label-count distinction holds in two noncyclic models (Appendix G, Figure 6).

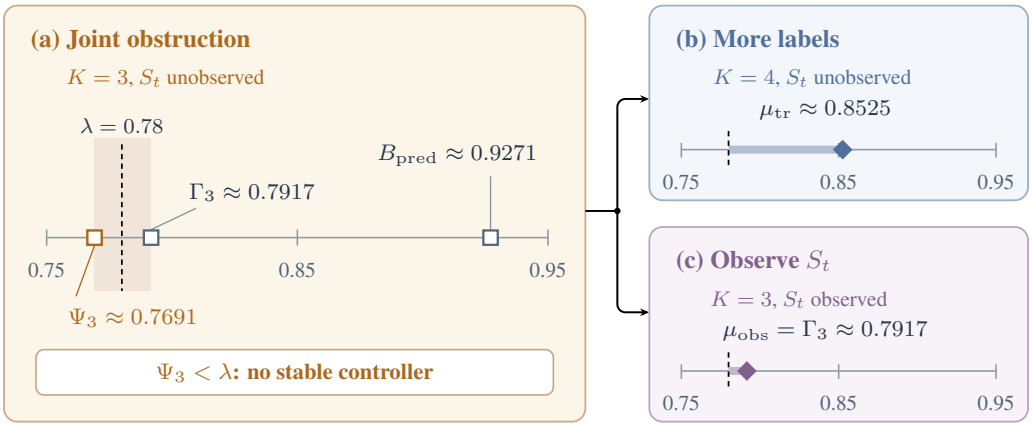

Figure 2: **MINT exclusion and two stabilizing interventions.** All panels use $m = 4$, $b = 1/4$, $g = 35/36$, $p = 15/16$ and arrival rate $\lambda = 0.78$ (dashed lines); rates are potential service per slot. **(a)** Open squares are upper bounds: joint Ψ_3; interface-only Γ_3, granting current-state information; and prediction-only B_{pred}, granting previous-state information and unrestricted labels. Only Ψ_3 excludes all compliant three-label departure-history controllers at this load in (Ψ_3, Γ_3). **(b,c)** A fourth label under unchanged feedback, or current-state observation with three labels, permits stable control. Filled diamonds are attained saturated rates; positive recurrence follows from the closed-loop proofs in Theorem 4.

Using MINT before training. Given $(\mathsf{H}, R, K, \lambda)$ satisfying the model assumptions, compute the invariant mode law π, form $\mathsf{H}R$, and enumerate menus to obtain $\Psi_K = \Gamma_K(\pi, \mathsf{H}R)$. A certified upper bound on this maximum also suffices: if it is below λ, the interface is infeasible. A low-value candidate menu alone cannot establish this verdict. In Figure 2, the separate bounds 0.7917 and 0.9271 exceed demand 0.78, whereas MINT gives 0.7691. Our constructions prove that adding a fourth command or supplying the current phase restores feasibility in this instance. This calculation assumes the specified model; an unqualified parameter estimate is not a certificate for an unknown environment.

Other loads and robustness. At higher loads than the joint-intervention interval, an interface-only bound also excludes $K < m$ while the tracker has a quantitative stability margin and exponential queue tails. Appendix D gives the separate theorem, perturbation certificates and Figure 4.

4 Closed-loop interventions and learned controllers

All deployment evaluations use actual departure feedback and continuing reflected queues without resets. We compare the command budgets excluded by MINT with closed-loop workloads, then ask whether reward-only learning can find controllers at a feasible budget.

4.1 Closed-loop interventions over the certified interval

We evaluate four loads in $\mathcal{I}_4$ and run each fixed controller for 100,000 slots on 32 paired environment seeds per load. Starts have empty queues, controller state zero, and uniform hidden phase; all loads and iid-phase controls appear in Appendix F.

Controller memory C_t has M states, distinct from the K labels. A three-state finite-state controller (FSC) and a four-state/three-label example both update $C_{t+1} = C_t + 1 - D_t$ modulo M; their action maps are $(0, 1, 2)$ and $(0, 1, 2, 2)$. The label intervention changes the latter map to $(0, 1, 2, 3)$; the observation intervention supplies current S_t and uses $(0, 1, 2, 0)$. Table 1 keeps these contracts separate.

Table 1: Closed-loop interventions at load 0.78: $T = 100{,}000$, 32 seeds. μ is model-solved; Q_T entries give means and 95% t-interval halfwidths, not recurrence classifications.

Controller / runtime input	M	K	Solved μ	Finite-run Q_T
Selected FSC / departures	3	3	0.7172	6222 ± 154
4-state example / departures	4	3	0.7099	6948 ± 160
More labels / departures	4	4	0.8525	5.2 ± 1.7
Current-phase observation	1	3	0.7917	59 ± 22

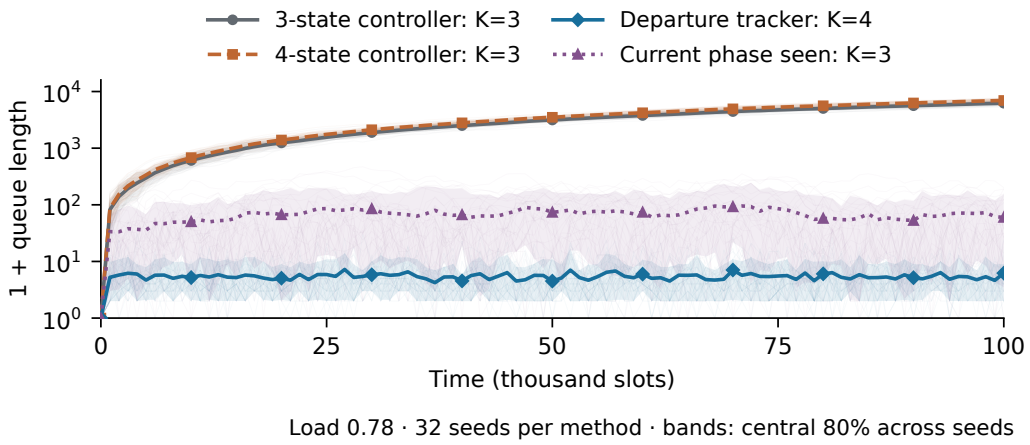

Figure 3: **Fixed-policy workload at load** 0.78**.** Faint lines show all 32 seeds; bold lines are arithmetic means, and bands are central 80% seed ranges, not confidence intervals. The ordinate is $1 + Q_t$ on a log scale. The four-state example also changes the update and action map relative to the selected FSC, so these curves alone do not isolate a memory effect.

In these fixed-policy runs across all four loads, both three-label history controllers accumulate backlog, while the four-label tracker stays near empty. The stable current-phase controller has larger queues near its service boundary; Appendix F gives the full grid.

4.2 When feedback cannot predict the next phase, more labels do not help

In an iid-phase variant, the next phase is independent of history. Every history-only action then matches with probability $1/4$; even unlimited labels cannot stabilize the main loads. The current-phase menu is stable by a separate recurrence proof. The interpolation formula and workload comparisons are in Appendix F.

4.3 Reward-only joint learning with an adequate interface

We study finite-controller search and recurrent neural learning in the same queue. The finite search uses $M = 4$ memory states and learns label choice, label-only action distributions, and departure-based memory updates from reward, without state, queue, or teacher targets. Training starts with enough backlog to estimate saturated service; frozen controllers are then evaluated from empty queues, with finite-controller stability checked separately (Appendix H).

Certified finite controllers. Reward-ranked mutation searches deterministic tables from random starts, without a tracker template or model access during search. Ten independent training seeds per interface use equal budgets; the search procedure is in Appendix H.

All ten four-label seeds recover the tracker up to renaming. A post-training verifier checks positive-queue drift and the empty boundary, proving recurrence and exponential queue moments. A $K = 8$ control also recovers the tracker; the three-label theorem excludes even randomized decoders. Appendix H reports every seed.

Table 2: Jointly learned controllers at $\lambda = 0.78$, $M = 4$ (ten seeds per row). "Certified" counts post-training stability proofs; μ is model-solved. $\overline{Q}_{\text{late}}$ averages the last half of million-slot empty-start queues; brackets give 95% bootstrap intervals that resample training and environment seeds separately.

Labels K	Certified	Solved μ	$\overline{Q}_{\text{late}}$ [95% interval]
3	0/10	0.7172	46825 [46276, 47318]
4	10/10	0.8525	4.418 [4.369, 4.470]

Joint recurrent neural learning. GRU-PPO uses past actions and departures to select command labels; Full uses the same history to select among four physical actions directly. Four labels can represent Full's action head, so their differences reflect training rather than action capacity.

With an auxiliary head predicting the observed chosen-action departure, raising only the four-label decoder's learning rate tenfold changes the 100,000-slot empirical pass count from 0/10 to 10/10. The corresponding three-label setup gives 0/10; ordinary unrestricted PPO reaches 10/10 with twice the update budget. These finite-run outcomes separate interface feasibility from training success. Appendix I retains all six conditions, exact gate criteria, conditional million-slot checks, stronger baselines, and nearby-environment readouts.

5 Related work and scope

Information and stability. Asanjarani & Nazarathy (2020) study Markov-modulated servers under partial observations, including service feedback, using belief controllers and queue-stability analysis. Our resource is the fixed label-only action menu, with unrestricted upstream memory. The new separation combines a class-wide joint exclusion with matching stable controllers in the same environment, establishing exact minimum command counts under two observation contracts. Other limits concern communication rate (Tatikonda & Mitter, 2004; Nair & Evans, 2004), sensor information (Majumdar et al., 2023), and dispatching memory/messages (Gamarnik et al., 2022).

Finite messages and policy representations. The posterior-gain form of Alvim et al. (2012) supplies the per-message decision step in (8); the fresh-transition conditional expectation supplies $\mathsf{H}R$. Our queue construction makes the resulting joint obstruction attain an exact command threshold. Unlike a history-reading decoder for reconstruction (Wood et al., 2017), ours reads one current label. Quantized policies and finite-state controllers study other finite representations (Saldi et al., 2015; Meuleau et al., 1999a; Andriushchenko et al., 2022). Recurrent RL uses history in POMDPs (Ni et al., 2022), and discrete multi-agent messages may reach privately informed receivers (Foerster et al., 2016); neither resource is given to our executor. Learning work (Meuleau et al., 1999b; Pavse et al., 2024) motivates our task-specific certified witnesses.

Limitations. The menu bound requires label-only execution; MINT additionally requires hidden-mode changes that remain fresh at decision time. Giving the executor current-mode observations, history-dependent templates, or action-visible memory changes the information interface. When MINT's ceiling exceeds arrivals, it makes no stability claim; a separate construction is needed. The exact one-command saving holds for the matched-cycle load interval and two certified noncyclic instances; the cost of current-mode sensing is outside our model. Neural rollouts are finite-run checks, not recurrence certificates or general algorithm comparisons. Nearby-task stochastic deployment and zero-shot transfer were unreliable (Appendix I).

6 Conclusion

MINT certifies growing expected backlog for every compliant history controller when limited commands and unseen mode changes jointly obstruct service. Matching stable constructions establish exact budgets of m commands with departure feedback and $m - 1$ with current-mode observation in the matched-cycle family. Two noncyclic instances share this separation; reward-only search recovers certified stable controllers at a feasible budget.

Reproducibility Statement

The paper specifies the observation and no-bypass execution contracts, event order, and model parameters. The appendices provide converse proofs, closed-loop drift and boundary arguments, finite-controller certification criteria, evaluation protocols, seed identifiers, and reported results. The supplementary archive contains manuscript source in `source/`, exact off-cycle certificates in `certificate/`, and computational evidence in `evidence/`. The evidence directory contains the finite-controller trainer, selection and evaluation code, configurations, initial populations, compact search histories, final controllers and exact certificates. Its entry point audits those records, rebuilds queue statistics, and replays frozen-controller evaluations; training requires an explicit mode. It also contains intervention tables and sampled traces, and derived neural summaries with a per-seed table generator. Neural checkpoints, full trajectories and training code are not included, so neural deployment cannot be reproduced from this archive alone. `evidence/CLAIM_EVIDENCE.md` maps finite-check claims to source, stored results and replay commands. Learned controllers are frozen before confirmation queue evaluations. Training seeds are the units of replication; environment trajectories and checkpoint readouts are not additional training runs. Neural empirical gates are distinct from exact finite-controller certificates.

AI Use Statement

Generative AI tools assisted research ideation, conceptual framing and hypothesis refinement; mathematical claims, proof development and checking; experimental design, implementation, analysis and tests; literature search and summarization; writing, captions and figures; and interpretation of results. We did not use generative AI to generate synthetic data sets. Verification includes executable tests, exact-arithmetic checks, citation checks and seed-level evidence audits, with the reproducibility limits stated above. Authors take responsibility for the submitted claims and artifacts.

References

Mário S. Alvim, Konstantinos Chatzikokolakis, Catuscia Palamidessi, and Geoffrey Smith. Measuring information leakage using generalized gain functions. In *2012 IEEE 25th Computer Security Foundations Symposium*, pp. 265–279, 2012. doi: 10.1109/CSF.2012.26.

Roman Andriushchenko, Milan Češka, Sebastian Junges, and Joost-Pieter Katoen. Inductive synthesis of finite-state controllers for POMDPs. In *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence*, volume 180 of *Proceedings of Machine Learning Research*, pp. 85–95. PMLR, 2022. URL https://proceedings.mlr.press/v180/andriushchenko22a.html.

Azam Asanjarani and Yoni Nazarathy. The role of information in system stability with partially observable servers. *Methodology and Computing in Applied Probability*, 22(3):949–968, 2020. doi: 10.1007/s11009-019-09750-4.

Jakob Foerster, Ioannis Alexandros Assael, Nando de Freitas, and Shimon Whiteson. Learning to communicate with deep multi-agent reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 29, 2016. URL https://proceedings.neurips.cc/paper/2016/hash/c7635bfd99248a2cdef8249ef7bfbef4-Abstract.html.

David Gamarnik, John N. Tsitsiklis, and Martin Zubeldia. Stability, memory, and messaging tradeoffs in heterogeneous service systems. *Mathematics of Operations Research*, 47(3):1862–1874, 2022. doi: 10.1287/moor.2021.1191.

Guy Latouche and Vaidyanathan Ramaswami. *Introduction to Matrix Analytic Methods in Stochastic Modeling*. ASA-SIAM Series on Statistics and Applied Probability. Society for Industrial and Applied Mathematics, 1999. ISBN 9780898714258. doi: 10.1137/1.9780898719734.

Anirudha Majumdar, Zhiting Mei, and Vincent Pacelli. Fundamental limits for sensor-based robot control. *The International Journal of Robotics Research*, 42(12):1051–1069, 2023. doi: 10.1177/02783649231190947.

Nicolas Meuleau, Kee-Eung Kim, Leslie Pack Kaelbling, and Anthony R. Cassandra. Solving POMDPs by searching the space of finite policies. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pp. 417–426, 1999a. URL https://dblp.org/rec/conf/uai/MeuleauKKC99.

Nicolas Meuleau, Leonid Peshkin, Kee-Eung Kim, and Leslie Pack Kaelbling. Learning finite-state controllers for partially observable environments. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pp. 427–436, 1999b. URL https://arxiv.org/abs/1301.6721.

Sean P. Meyn and Richard L. Tweedie. Stability of Markovian processes I: Criteria for discrete-time chains. *Advances in Applied Probability*, 24(3):542–574, 1992. doi: 10.2307/1427479.

Sean P. Meyn and Richard L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, 2nd edition, 2009. ISBN 9780521731829. doi: 10.1017/CBO9780511626630.

Girish N. Nair and Robin J. Evans. Stabilizability of stochastic linear systems with finite feedback data rates. *SIAM Journal on Control and Optimization*, 43(2):413–436, 2004. doi: 10.1137/S0363012902402116.

Tianwei Ni, Benjamin Eysenbach, and Ruslan Salakhutdinov. Recurrent model-free RL can be a strong baseline for many POMDPs. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 16691–16723. PMLR, 2022. URL https://proceedings.mlr.press/v162/ni22a.html.

Brahma S. Pavse, Matthew Zurek, Yudong Chen, Qiaomin Xie, and Josiah P. Hanna. Learning to stabilize online reinforcement learning in unbounded state spaces. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 40015–40033. PMLR, 2024. URL https://proceedings.mlr.press/v235/pavse24a.html.

Naci Saldi, Tamás Linder, and Serdar Yüksel. Asymptotic optimality and rates of convergence of quantized stationary policies in stochastic control. *IEEE Transactions on Automatic Control*, 60 (2):553–558, 2015. doi: 10.1109/TAC.2014.2343831. URL https://doi.org/10.1109/TAC.2014.2343831.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017. URL https://arxiv.org/abs/1707.06347.

Paul J. Schweitzer. Perturbation theory and finite Markov chains. *Journal of Applied Probability*, 5 (2):401–413, 1968. doi: 10.2307/3212261.

Sekhar Tatikonda and Sanjoy Mitter. Control under communication constraints. *IEEE Transactions on Automatic Control*, 49(7):1056–1068, 2004. doi: 10.1109/TAC.2004.831187.

Ronald J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8:229–256, 1992. doi: 10.1007/BF00992696.

Richard G. Wood, Tamás Linder, and Serdar Yüksel. Optimal zero delay coding of Markov sources: Stationary and finite memory codes. *IEEE Transactions on Information Theory*, 63(9):5968–5980, 2017. doi: 10.1109/TIT.2017.2692215.

A Proof of Theorem 1 and Corollary 7

Fix t and write $\xi_t(s,z) = \Pr(S_t = s, Z_t = z)$. The structural factorization makes the action kernel $q_t(\cdot \mid z)$ independent of s after z is fixed. Hence

$$\mathbb{E}\widetilde{J}_t = \sum_{s,z} \xi_t(s,z) \sum_a q_t(a \mid z) R(s,a) \tag{18}$$

$$\leq \sum_z \max_a \sum_s \xi_t(s,z) R(s,a). \tag{19}$$

For each non-null label choose a maximizing action a_z; assign an arbitrary action to null labels. Then

$$\sum_z \max_a \sum_s \xi_t(s,z) R(s,a) = \sum_{s,z} \xi_t(s,z) R(s,a_z) \tag{20}$$

$$\leq \sum_s \nu_t(s) \max_{z \leq K} R(s,a_z) \leq \Gamma_K(\nu_t, R), \tag{21}$$

which proves (4) without any independence assumption on (S_t, Z_t).

Since $D_t \leq \widetilde{J}_t$, (1) and the preceding bound give

$$\mathbb{E}(Q_{t+1} - Q_t) \geq \lambda - \Gamma_K(\nu_t, R).$$

Summing from $t = 0$ to $T-1$ proves (5); the assumption $\mathbb{E}Q_0 < \infty$ and bounded one-step increments give integrability at each finite T.

For a fixed menu $M = (a_1, \ldots, a_K)$, put $f_M(s) = \max_j R(s, a_j)$. Its oscillation is at most $r_{\max} - r_{\min}$. The finite-space total-variation inequality therefore gives

$$\sum_s \nu(s) f_M(s) \leq \sum_s \pi(s) f_M(s) + (r_{\max} - r_{\min})\, \mathrm{TV}(\nu, \pi).$$

Maximizing the left side over M, and bounding the π expectation of each such menu by $\Gamma_K(\pi, R)$, proves (6).

For invariant-law exclusion, suppose a time-homogeneous full augmented realization had invariant distribution Π. Autonomy implies that its S marginal is invariant for the hidden kernel and hence equals the unique law π. Let $f_N(q) = \min(q, N)$. Stationarity gives $\mathbb{E}_\Pi[f_N(Q_{t+1}) - f_N(Q_t)] = 0$. For every realized transition the increment converges to $Q_{t+1} - Q_t = X_t - D_t$ as $N \to \infty$ and has absolute value at most one. Dominated convergence yields $0 = \lambda - \mathbb{E}_\Pi D_t$. But $D_t \leq \widetilde{J}_t$ and the already-proved menu bound under hidden marginal π give $\mathbb{E}_\Pi D_t \leq \Gamma_K(\pi, R) < \lambda$, a contradiction.

Menu number and representation budgets. For arrival rate λ, define the minimum capacity-clearing menu number

$$K^\star(\lambda, \nu, R) = \min\{K : \Gamma_K(\nu, R) > \lambda\}, \tag{22}$$

with $K^\star = \infty$ if no finite menu clears the rate. This number measures *strict-slack menu clearance*. The converse excludes $\lambda > \Gamma_K$; critical equality is not classified in general. Menu clearance also does not imply that an observation-constrained controller can attain the optimizing menu.

Proposition 5 (Exact current-phase menu threshold)**.** *Suppose* H *is finite and irreducible,* $0 < \lambda < 1$, $0 < r_{\min} \leq R(s,a) \leq r_{\max} < 1$*, and, conditional on the full predecision past and* (Q_t, S_t, A_t)*, the arrival, potential service, and next phase are independent fresh draws with parameters* λ*,* $R(s,a)$*, and* $\mathsf{H}(s,\cdot)$*. Keep a fixed, label-only decoder and let the encoder observe* S_t*. A positive-recurrent homogeneous* K*-label controller then exists if and only if* $\lambda < \Gamma_K(\pi, R)$*, where* π *is the invariant phase law. Consequently its minimum stable label count is* $K^{\min}_{\text{phase}} = K^\star(\lambda, \pi, R)$*, with both sides equal to* ∞ *if no finite menu clears demand.*

Proof. If $\mu := \Gamma_K(\pi, R) > \lambda$, choose an optimizing menu of at most K physical actions and, in phase s, send the label of an action maximizing $R(s,a)$ within that menu. The resulting service mean $r(s)$ satisfies $\pi r = \mu$. Finite irreducibility gives a bounded solution to $(I - \mathsf{H})h = \mu\mathbf{1} - r$. For $V(q,s) = q + h(s) - \min_u h(u)$ and $q > 0$, the fresh primitives give $\mathbb{E}[\Delta V \mid q, s] = \lambda - \mu < 0$. Drift at $q = 0$ is bounded. All arrival and service outcomes have positive probability, and the phase

chain is irreducible, so the reachable (Q, S) chain is irreducible. The Foster criterion yields positive recurrence.

Conversely, let a compliant homogeneous augmented controller have an invariant law. Truncating Q and using bounded increments as in the proof of Theorem 1 gives $\lambda = \mathbb{E}D_t$. Its invariant phase marginal is π, so $\mathbb{E}\widetilde{J}_t \le \Gamma_K(\pi, R)$. The invariant law must have $\Pr(Q_t = 0) > 0$: otherwise take the least queue value with positive invariant mass; from that positive value, independent no arrival and successful service move down with probability at least $(1-\lambda)r_{\min} > 0$, contradicting least support. Therefore

$$\lambda = \mathbb{E}D_t = \mathbb{E}\widetilde{J}_t - \mathbb{E}[\mathbf{1}\{Q_t = 0\}\widetilde{J}_t] \le \Gamma_K(\pi, R) - r_{\min}\Pr(Q_t = 0) < \Gamma_K(\pi, R).$$

This also excludes critical equality under the proposition's stronger primitive assumptions. □

This exact characterization concerns current-phase observation. Under departure-only feedback, Ψ_K grants the encoder the previous phase as an oracle and need not be attainable; neither $\Psi_K > \lambda$ nor $\Gamma_K > \lambda$ proves history-only stability.

Alphabet and bit budgets. Menu inclusion gives $\Gamma_{K_1} \le \Gamma_{K_2}$ for $K_1 \le K_2$. A decoder reading only a fixed-width B-bit word has at most 2^B rows, so its bound is Γ_{2^B}. On the interval of Theorem 4, the minimum widths are $\lceil \log_2 m \rceil$ under departure feedback and $\lceil \log_2(m-1) \rceil$ with current-phase observation. They can be equal (for example, at $m = 4$); saving one label need not save one bit. These are instantaneous message widths and exclude the cost of observing the phase.

Proposition 6 (Hardness of the capacity-clearing menu)**.** *Given explicitly encoded rational (ν, R, λ) and a positive integer K, deciding whether $K^\star(\lambda, \nu, R) \le K$ is NP-complete. This remains true when ν is uniform, $R(s,a) \in \{0,1\}$, and $\lambda = 1 - 1/(2|\mathcal{S}|)$.*

We now prove Proposition 6. Membership in NP follows because repetitions are unnecessary: a certificate lists at most $\min(K, |\mathcal{A}|)$ distinct action indices, even when K is binary encoded. For an explicitly given rational matrix and law, its row maxima, weighted sum, and comparison with λ have polynomial bit complexity. For NP-hardness, reduce from SET COVER. Given universe $\mathcal{S}$ of size N and sets $(B_a)_{a\in\mathcal{A}}$, let $\nu = u_N$, set $R(s,a) = \mathbf{1}\{s \in B_a\}$, and choose $\lambda = 1 - 1/(2N)$. For any menu M,

$$\sum_s u_N(s) \max_{a\in M} R(s,a) = \frac{|\bigcup_{a\in M} B_a|}{N}.$$

This value is an integer multiple of $1/N$ and exceeds $1 - 1/(2N)$ if and only if it equals one. Hence $K^\star \le K$ exactly when the original instance has a cover using at most K sets. The construction is polynomial and satisfies all restrictions in the proposition.

Computational implication for the current-phase queue. Under Proposition 5's assumptions, deciding whether a K-label current-phase controller stabilizes the queue is also NP-complete for explicit rational inputs. Membership follows from an action menu certificate, the proposition, and rational linear algebra for the invariant law of H. For hardness, in the preceding SET COVER reduction take the iid uniform phase kernel and replace the binary service matrix C by $R = 1/4 + C/2$, with $\lambda = 3/4 - 1/(4N)$. Then all service probabilities are strictly between zero and one, and $\Gamma_K(\pi, R) > \lambda$ exactly when the selected sets cover the universe. This is finite enumeration, not a polynomial-time algorithm for the minimum budget; it does not characterize the departure-feedback minimum.

It remains to specialize. Under u_m, a menu containing d distinct matched actions obtains reward g on exactly d phases and b on the other $m - d$. Thus $d \le \min(K, m)$ gives $\Gamma_K(u_m, R) = b + (g-b)\min(K,m)/m$, with equality for any menu containing $\min(K,m)$ distinct actions. The hidden cycle is irreducible and doubly stochastic, so u_m is its unique invariant law and is preserved from a uniform start. Applying Theorem 1 proves (24); applying (6) with service range $g - b$ proves (25). Finite irreducibility and aperiodicity give geometric convergence to u_m, so the TV series in (26) is finite. The invariant-law conclusion proves the last assertion of Corollary 7.

B MATCHED-CYCLE CAPACITY AND CONSTRUCTIVE WITNESS

Corollary 7 (Exact matched-cycle menu capacity). *For uniform u_m,*

$$\Gamma_K(u_m, R) = c_K := b + (g-b)\frac{\min(K,m)}{m}. \tag{23}$$

The menu optimization is attained by any menu containing $\min(K,m)$ distinct actions; this is not a claim that the observation-constrained controller attains the menu value at every time. If $S_0 \sim u_m$, every no-bypass controller satisfies

$$\mathbb{E}Q_T \geq \mathbb{E}Q_0 + (\lambda - c_K)T. \tag{24}$$

For arbitrary S_0,

$$\mathbb{E}Q_T \geq \mathbb{E}Q_0 + (\lambda - c_K)T - (g-b)\sum_{t<T} \mathrm{TV}(\nu_t, u_m). \tag{25}$$

Because (13) is finite, irreducible, and aperiodic,

$$C_m(\nu_0) := (g-b)\sum_{t\geq 0} \mathrm{TV}(\nu_t, u_m) < \infty, \tag{26}$$

so (25) is at least $\mathbb{E}Q_0 + (\lambda - c_K)T - C_m(\nu_0)$. If $\lambda > c_K$, every time-homogeneous augmented realization has no invariant probability law.

Remark 8 (Scope). *The queue bound is in expectation; it does not assert almost-sure escape for arbitrary nonstationary encoders. Autonomy identifies the stationary hidden marginal in the invariant-law conclusion, while the pointwise menu inequality uses the actual hidden marginal.*

Let $n = m - 1$ and choose

$$b = \tfrac{1}{4}, \quad g = 1 - \tfrac{1}{4n^2}, \quad p = 1 - \tfrac{1}{m^2}, \quad \Delta_m = \tfrac{13n}{320m^2}, \quad \lambda_m = c_K\big|_{K=m-1} + \Delta_m. \tag{27}$$

Theorem 9 (Rational matched-observation separation). *For every integer $m \geq 2$, parameters (27) satisfy:*

1. *For every initial hidden law ν_0 and $\mathbb{E}Q_0 < \infty$, every $K \leq m-1$ no-bypass factorization of (2) obeys $\mathbb{E}Q_T \geq \mathbb{E}Q_0 + \Delta_m T - C_m(\nu_0)$ for all $T \geq 0$; for $\nu_0 = u_m$, the constant is zero.*
2. *The m-label outcome tracker (17) has a positive recurrent irreducible communicating class containing the initialization in the augmented chain (Q_t, S_t, A_t). Its saturated phase chain has stationary potential-service rate μ_m satisfying $\mu_m - \lambda_m > \Delta_m > 0$.*

The first part is Corollary 7 at the $m-1$ ceiling. For the second, let $\Phi_t = S_t - A_t \pmod m$ at positive queue levels. Write $\theta_0 = g$ and $\theta_d = b$ for $d \neq 0$. Its saturated transition kernel is

$$\begin{aligned} P(d,e) = \theta_d\big[p\mathbf{1}\{e=d\} + (1-p)\mathbf{1}\{e=d+1\}\big] \\ + (1-\theta_d)\big[p\mathbf{1}\{e=d-1\} + (1-p)\mathbf{1}\{e=d\}\big], \end{aligned} \tag{28}$$

where all indices are modulo m. Thus parallel events are summed, including $e = d+1 = d-1$ when $m = 2$. Solving its stationary current recurrence gives a lower bound on match mass,

$$\pi_0 > \frac{n}{n+1} + \frac{13n}{80(n+1)^2}. \tag{29}$$

Here μ_m is computed in the saturated phase chain, not by conditioning the stationary reflected queue on being nonempty. Thus $\mu_m = b + (g-b)\pi_0$ exceeds the $m-1$ ceiling by more than $2\Delta_m$.

We establish recurrence using a phase-dependent drift correction. Let $I_t = (S_t, A_t)$ be the finite positive-level phase and P° its phase kernel.

Lemma 10 (Communication of the tracker chain). *For the rational parameters in (27), the finite positive-level phase kernel P° is irreducible and aperiodic. The reachable full-state class of (Q_t, S_t, A_t) communicates between its finite level-zero boundary and every positive queue level.*

At fixed $q \geq 1$, success–arrival–hidden-increment events generate $(S, A) \mapsto (S+1, A)$, while failure–no-arrival–hidden-hold events generate $(S, A) \mapsto (S, A+1)$; both have strictly positive probability. These two increments generate $\mathbb{Z}_m^2$. A success with hidden hold gives a positive self-loop in P°. Failure with arrival raises the queue, success without arrival lowers it, and fixed-level paths repair the phase. An arrival leaves the finite boundary, while a success without arrival returns from level one. Appendix C supplies the event-by-event argument, including $m = 2$. If $d(i) = \mathbb{E}[Q_{t+1} - Q_t \mid Q_t \geq 1, I_t = i]$, its stationary mean is $\bar{d} = \lambda_m - \mu_m < 0$. Solve the finite Poisson equation $h - P^\circ h = d - \bar{d}\mathbf{1}$. Then

$$V(q, i) = q + h(i) - \min_j h(j) \tag{30}$$

is at least q and, for every $q \geq 1$,

$$\mathbb{E}[V(Q_{t+1}, I_{t+1}) - V(q, i) \mid Q_t = q, I_t = i] = d(i) + (P^\circ h)(i) - h(i) = \bar{d} < 0. \tag{31}$$

Lemma 10 makes the zero queue a finite communicating boundary, so the countable-state Foster–Lyapunov criterion (Meyn & Tweedie, 1992; 2009) proves positive recurrence. This is a finite-phase QBD structurally (Latouche & Ramaswami, 1999), but the proof uses the explicit Lyapunov bridge; Appendix C contains the current algebra and boundary argument.

C Proof details for Theorem 9

The undersized-label conclusion first requires that λ_m be a valid Bernoulli mean. Because

$$1 - c_{m-1} = \frac{3/4 + 1/(4n)}{n+1} > \frac{13n}{320(n+1)^2} = \Delta_m, \tag{32}$$

we have $0 < \lambda_m = c_{m-1} + \Delta_m < 1$. For every $K \leq m-1$, $c_K \leq c_{m-1}$, so Corollary 7 gives, from any hidden initial law,

$$\mathbb{E}Q_T \geq \mathbb{E}Q_0 + (\lambda_m - c_K)T - C_m(\nu_0) \geq \mathbb{E}Q_0 + \Delta_m T - C_m(\nu_0). \tag{33}$$

It remains to show the tracker's service margin and recurrence. Write $q = 1 - p = 1/m^2$ and let π be the stationary law of (28). With

$$r = \frac{bq}{(1-b)p} = \frac{1}{3n(n+2)}, \quad B = \frac{1-g}{1-b} = \frac{1}{3n^2}, \quad C = \frac{B - 4gr^{n+1}}{1 - r^{n+1}}, \tag{34}$$

the stationary-current recurrence has solution

$$\frac{\pi_i}{\pi_0} = C + (4g - C)r^i, \quad 1 \leq i \leq n. \tag{35}$$

For $m = 2$, the two directed moves share a destination but remain parallel events, so their aggregate satisfies the same current equation. Indeed, the current equations are $\pi_1/\pi_0 = 4gr - j$, $\pi_{i+1}/\pi_0 = r\pi_i/\pi_0 - j$, and $j = r\pi_n/\pi_0 - B$; solving them yields (35). The numerator of C is positive because $r \leq 1/9$ and $4gr^{n+1} < 4r^2 \leq 4/(81n^2) < B$. Summing (35) gives

$$n(\pi_0^{-1} - 1) = n^2 C + n(4g - C)\frac{r(1 - r^n)}{1 - r}. \tag{36}$$

Since $r^{n+1} \leq r^2 \leq 1/81$, the first term is less than $1/[3(1 - r^{n+1})] \leq 27/80$. Since $C > 0$, $g < 1$, and $r \leq 1/9$, the second is less than $4nr/(1-r) \leq 1/2$. Hence it is less than $67/80$. Therefore

$$\pi_0 > \frac{80n}{80n + 67} > \frac{n}{n+1} + \frac{13n}{80(n+1)^2}. \tag{37}$$

As $g - b \geq \frac{1}{2}$, this implies $\mu_m - c_{m-1} > 13n/(160m^2) = 2\Delta_m$, hence $\mu_m - \lambda_m > \Delta_m$.

For the recurrence step, let $I_t = (S_t, A_t)$ be the positive-level phase, P° its kernel after summing over queue increments, and $d(i)$ the conditional level drift. We first prove Lemma 10. At any fixed $q \geq 1$ and phase (s, a), the event

$$\{\widetilde{J}_t = 1,\ X_t = 1,\ H_t = 1\}$$

keeps the queue fixed and maps (s, a) to $(s + 1, a)$. Its probability is $R(s, a)\lambda_m(1 - p) > 0$. The event

$$\{\widetilde{J}_t = 0,\ X_t = 0,\ H_t = 0\}$$

also keeps the queue fixed and maps (s, a) to $(s, a+1)$; its probability is $(1-R(s, a))(1-\lambda_m)p > 0$. These coordinate increments generate $\mathbb{Z}_m^2$, proving irreducibility of P°. A success and hidden hold leave both phase coordinates unchanged, so P° has a self-loop of probability at least $bp > 0$ and is aperiodic. This argument also covers $m = 2$: the two generators remain distinct positive-probability event paths even when some difference-chain edges have the same endpoint.

Failure with arrival increases the queue by one, and success without arrival decreases it by one; both events have positive probability from every positive phase. After either move, the fixed-level generators can reach any desired phase. At $q = 0$, actual departure is zero, so an arrival enters level one with positive probability. Conversely, from level one a success without an arrival returns to level zero. The level-zero phases communicate as well: with no arrivals, m consecutive zero actual departures return A to its starting value, and choosing hidden increments within those steps generates the hidden-cycle coordinate; single zero-departure–hidden-hold steps generate the action coordinate. Thus the reachable full-state class is irreducible, its level-zero boundary is finite, and it communicates with every positive level. This proves the lemma.

Under the unique stationary law η of P°, its pushforward by $(s, a) \mapsto s - a \pmod m$ is stationary for (28); uniqueness therefore identifies that pushforward with π. Consequently the mean drift is $\bar{d} = \eta d = \lambda_m - \mu_m < 0$. Because P° is a finite irreducible chain, the centered vector $d - \bar{d}\mathbf{1}$ lies in the range of $I - P^\circ$, so there exists a finite solution of $h - P^\circ h = d - \bar{d}\mathbf{1}$. Thus $V(q, i) = q + h(i) - \min_j h(j) \geq q$ has, for every $q \geq 1$,

$$\mathbb{E}[V(Q_{t+1}, I_{t+1}) - V(q, i) \mid Q_t = q, I_t = i] = d(i) + (P^\circ h)(i) - h(i) = \bar{d} < 0. \tag{38}$$

The finite sublevel sets of V and the negative drift on every positive level therefore meet the hypotheses of the countable-state Foster criterion (Meyn & Tweedie, 1992; 2009), which proves positive recurrence.

D Quantitative robustness and predictability

This appendix proves Proposition 3 and the higher-load results summarized in Section 3.3, including exponential queue tails and tolerance to model perturbations.

D.1 Matched-cycle model

Let $m \geq 2$, $n = m - 1$, and $S_t, A_t \in \mathbb{Z}_m$. The hidden process obeys

$$S_{t+1} = S_t + H_t \pmod m, \qquad \Pr(H_t = 0) = p, \quad \Pr(H_t = 1) = 1 - p.$$

Potential service has conditional mean

$$R(s, a) = b + (g - b)\mathbf{1}\{s = a\}, \qquad 0 < b < g < 1.$$

The queue and actual departures satisfy

$$D_t = \mathbf{1}\{Q_t > 0\}\widetilde{J}_t, \qquad Q_{t+1} = Q_t - D_t + X_t.$$

Conditional on the predecision past and the current state and action, $\widetilde{J}_t$, X_t, and H_t are mutually independent Bernoulli draws, with means $R(S_t, A_t)$, λ, and $1 - p$. The event order is action, service, departure, arrival, hidden transition. The controller sees actual departures, not the current hidden phase or queue. In particular, the innovation H_{t-1} is not observed before action A_t.

A no-bypass K-label action law has a label $Z_t \in [K]$ and decoder rows $q_t(\cdot \mid z)$, with action randomness screened off from all other current information by the label. The interface-only menu theorem gives

$$\mathbb{E}\widetilde{J}_t \leq \Gamma_K(\nu_t, R), \qquad \Gamma_K(\nu, R) = \max_{a_1,\dots,a_K} \sum_s \nu(s) \max_{j \leq K} R(s, a_j). \tag{39}$$

The uniform hidden law u_m is invariant and

$$c_K = \Gamma_K(u_m, R) = b + (g-b)\frac{\min(K,m)}{m}. \tag{40}$$

The converse even permits the encoder to know S_t, provided the decoder still satisfies screening. All achievability and predictability statements below retain departure-only feedback.

The specified tracker is unchanged:

$$A_t = Z_t, \qquad Z_{t+1} = Z_t + 1 - D_t \pmod m. \tag{41}$$

At positive queue levels, its difference phase is $\Phi_t = S_t - Z_t \pmod m$. Write $\theta_0 = g$ and $\theta_d = b$ for $d \neq 0$. Its saturated kernel is

$$\begin{aligned} P(d,e) = \theta_d\big[p\mathbf{1}\{e=d\} + (1-p)\mathbf{1}\{e=d+1\}\big] \\ + (1-\theta_d)\big[p\mathbf{1}\{e=d-1\} + (1-p)\mathbf{1}\{e=d\}\big]. \end{aligned} \tag{42}$$

All indices are modulo m; parallel events are added, also when $m = 2$. Let π be its unique invariant law and $\mu = b + (g-b)\pi_0$ its stationary potential-service rate. This is not service conditioned on a reflected queue being nonempty.

D.2 An excursion certificate and an explicit robust family

Lemma 11 (Excursion service certificate)**.** *If $p > b$, the tracker satisfies*

$$\pi_0 \geq \frac{p-b}{p-b+g(1-p)+n(1-g)p}. \tag{43}$$

Consequently,

$$\mu \geq b + (g-b)\frac{p-b}{p-b+g(1-p)+n(1-g)p}. \tag{44}$$

Proof. Let $\tau_0 = \inf\{t \geq 0 : \Phi_t = 0\}$ when $\Phi_0 = i \neq 0$. On the representatives $0, 1, \ldots, n$, use $v(i) = i$. Before hitting zero and away from the top endpoint, its one-step drift is

$$b(1-p) - (1-b)p = b - p.$$

At $i = n$, an upward cyclic step lands at zero rather than $n+1$, so the drift is even smaller. This calculation also sums the two parallel moves at $m = 2$. Stopping at $\tau_0 \wedge T$ gives

$$0 \leq \mathbb{E}_i v(\Phi_{\tau_0 \wedge T}) \leq i - (p-b)\mathbb{E}_i(\tau_0 \wedge T).$$

Monotone convergence implies $\mathbb{E}_i\tau_0 \leq i/(p-b)$.

Starting at zero, the nonzero destinations in the first step are 1 with event probability $g(1-p)$ and n with event probability $(1-g)p$. They may coincide; their event contributions still add. The mean first return time is therefore

$$\mathbb{E}_0\tau_0^+ = 1 + g(1-p)\mathbb{E}_1\tau_0 + (1-g)p\mathbb{E}_n\tau_0 \leq 1 + \frac{g(1-p)+n(1-g)p}{p-b}.$$

The return-time identity $\pi_0 = (\mathbb{E}_0\tau_0^+)^{-1}$ proves the claim. □

Theorem 12 (Explicit scaled robust separation)**.** *For every integer $m \geq 2$, choose independently*

$$b, \kappa, \zeta \in [1/8, 1/4], \qquad \ell \in [1/32, 1/16], \tag{45}$$

and set

$$p = 1 - \frac{\kappa}{m}, \qquad g = 1 - \frac{\zeta}{(m-1)^2}, \qquad \lambda = c_{m-1} + \ell\frac{m-1}{m^2}. \tag{46}$$

These parameters are valid Bernoulli means. The following hold.

Undersized interfaces. *For every $K < m$ no-bypass controller, every initial hidden law ν_0, and $\mathbb{E}Q_0 < \infty$,*

$$\mathbb{E}Q_T \geq \mathbb{E}Q_0 + \frac{m-1}{32m^2}T - C(\nu_0), \qquad T \geq 0, \tag{47}$$

where $C(\nu_0) < \infty$ and $C(u_m) = 0$. Every compliant time-homogeneous full augmented realization has no invariant probability law.

The specified sufficient interface. *Tracker (41) has*

$$\mu - c_{m-1} > \frac{m-1}{8m^2}, \qquad \mu - \lambda > \frac{m-1}{16m^2}. \tag{48}$$

Its full queue–phase chain is irreducible and positive recurrent. Moreover, define

$$\delta_m = \frac{m-1}{16m^2}, \quad L_m = 1+\frac{6}{5}(m-1), \quad \vartheta_m = \frac{\delta_m}{3L_m^2}, \quad \rho_m = 1-\frac{\vartheta_m \delta_m}{2}, \quad B_m = \frac{e^{\vartheta_m L_m}}{1-\rho_m}. \tag{49}$$

For every deterministic finite $Q_0 = q_0$ and every initial phase,

$$\mathbb{E}e^{\vartheta_m Q_t} \le \rho_m^t e^{\vartheta_m (q_0 + 6(m-1)/5)} + B_m. \tag{50}$$

The invariant law Π satisfies

$$\Pi(Q \ge x) \le B_m e^{-\vartheta_m x}, \qquad x \ge 0. \tag{51}$$

Thus $\sup_t \mathbb{E}Q_t < \infty$ for these finite deterministic starts. More generally, (50) extends to random starts having a finite ϑ_m-exponential moment by averaging its initial term.

Proof. We have $p \ge 7/8$, $g \ge 3/4$, and $g - b \ge 1/2$. Also

$$\lambda - c_{m-1} \le \frac{n}{16m^2} < \frac{1}{16m} < \frac{g-b}{m} = g - c_{m-1},$$

so $0 < \lambda < g < 1$.

Match mass. By Lemma 11,

$$\begin{aligned} n(\pi_0^{-1} - 1) &\le \frac{ng(1-p) + n^2(1-g)p}{p-b} \\ &= \frac{ng\kappa/m + \zeta p}{p-b} < \frac{\kappa + \zeta p}{p-b} \le \frac{(1+p)/4}{p-b} \le \frac{3}{4}. \end{aligned} \tag{52}$$

For the last step, $(1+p)/4 \le 3(p-b)/4$ follows from $p \ge 7/8$ and $b \le 1/4$. The strict inequality uses $ng/m < 1$ and $\kappa > 0$. Hence

$$\pi_0 > \frac{4n}{4n+3}, \qquad \pi_0 - \frac{n}{n+1} > \frac{n}{(4n+3)(n+1)} > \frac{n}{4m^2}.$$

Multiplication by $g - b \ge 1/2$ gives the first inequality in (48); subtracting $\ell n/m^2 \le n/(16m^2)$ gives the second.

Converse and invariant-law exclusion. Equation (39) and $D_t \le \widetilde{J}_t$ imply

$$\mathbb{E}Q_T \ge \mathbb{E}Q_0 + (\lambda - c_K)T - (g-b)\sum_{t<T} \mathrm{TV}(\nu_t, u_m).$$

The finite hidden chain is irreducible and aperiodic, so the infinite total-variation sum is finite. Since $c_K \le c_{m-1}$, this proves (47). For a hypothetical invariant law of a homogeneous full augmented realization, the hidden marginal is u_m. Applying stationarity to $\min(Q, N)$ and using bounded one-step queue increments, dominated convergence yields $\mathbb{E}D_t = \lambda$ without requiring a finite stationary queue mean. But $\mathbb{E}D_t \le c_K < \lambda$, a contradiction. This statement concerns the full homogeneous realization, not an undefined invariant law for an arbitrary nonstationary policy.

Communication and the boundary. At a fixed level $q \ge 1$, the positive-probability event "success, arrival, hidden increment" changes (S, Z) to $(S+1, Z)$ while leaving q fixed. The event "failure, no arrival, hidden hold" changes it to $(S, Z+1)$ while leaving q fixed. These generate all of $\mathbb{Z}_m^2$. Failure with an arrival raises the level; success without an arrival lowers it. Every boundary state can enter level one, and every boundary state is reached from its corresponding level-one phase by success, no arrival, and hidden hold. Thus the full countable state space communicates. Success with an arrival and hidden hold gives a self-loop at positive levels.

A bounded Poisson correction. Normalize $h(0) = 0$ and set

$$h(i) = (\mu - b)\mathbb{E}_i\tau_0, \qquad i \neq 0.$$

The hitting-time equations give $(I - P)h = \mu - \theta$ off zero. At zero, the return-time identity gives

$$(Ph)(0) = (\mu - b)(\pi_0^{-1} - 1) = (g - b)(1 - \pi_0) = g - \mu,$$

so the same Poisson equation holds everywhere. Furthermore,

$$0 \leq h(i) \leq \frac{(g-b)n}{p-b} \leq \frac{(1-b)n}{p-b} \leq \frac{6}{5}n.$$

The last inequality follows from $5 + b \leq 6p$. Define $V(q, s, z) = q + h(s - z)$. At every positive level, the actual phase update equals the saturated one, hence

$$\mathbb{E}[\Delta V \mid q, s, z] = \lambda - \theta_{s-z} + (Ph)(s - z) - h(s - z) = \lambda - \mu < -\delta_m.$$

The function has finite sublevel sets, and the boundary is finite. The countable-state Foster criterion establishes positive recurrence, as in Appendix C (Meyn & Tweedie, 2009).

Exponential moments, not merely recurrence. At every level, $|\Delta V| \leq L_m$. At positive levels, Taylor's inequality gives, for $\vartheta > 0$,

$$\mathbb{E}e^{\vartheta\Delta V} \leq 1 - \vartheta\delta_m + \frac{1}{2}\vartheta^2 L_m^2 e^{\vartheta L_m}.$$

For $\vartheta = \vartheta_m$, we have $\vartheta_m L_m \leq 1$ and $e^{\vartheta_m L_m} < 3$. Thus the right side is at most ρ_m. Writing $W = e^{\vartheta_m V}$, a boundary transition has $V_{t+1} \leq L_m$, and therefore

$$\mathbb{E}[W_{t+1} \mid Q_t, S_t, Z_t] \leq \rho_m W_t + e^{\vartheta_m L_m}\mathbf{1}\{Q_t = 0\}.$$

Iteration and $W_t \geq e^{\vartheta_m Q_t}$ prove (50). The irreducible aperiodic positive recurrent chain converges from a point to Π. Apply this convergence first to bounded truncations of W, then use monotone convergence and the uniform moment bound, to obtain $\Pi W \leq B_m$. Markov's inequality proves (51). This proves the exponential-moment claim separately from positive recurrence. □

Proof of Proposition 3 for arbitrary strict-slack loads. The bound in (52), yielding (48), gives $\mu_{\mathrm{tr}} = \mu > c_{m-1}$ for every physical parameter choice in (15); the saturated phase kernel does not depend on λ. For any $0 < \lambda < \mu$, the same finite irreducible phase kernel and Poisson solution used above give $\mathbb{E}[\Delta V \mid q, s, z] = \lambda - \mu < 0$ at every $q \geq 1$. The communication argument applies because all primitive probabilities are strictly interior, so the finite boundary and Foster criterion give positive recurrence. Finally, ΔV is bounded. Taking a sufficiently small $\vartheta > 0$ in the Taylor estimate used above gives a geometric drift for $e^{\vartheta V}$ outside the finite boundary; its boundary expectation is finite. Consequently the invariant law has a finite exponential moment of Q, and hence an exponential queue tail. The explicit constants in Theorem 12 apply to its more restrictive higher-load family.

Remark 13 (What is and is not uniform)**.** *The box in (b, κ, ζ, ℓ) has fixed positive volume, independently of m. For each fixed m, its interior maps invertibly to an open region of physical parameters (b, p, g, λ). However, the physical widths in p, g, and λ shrink as m grows. Neither the physical volume, the service gap, nor the exponential-tail exponent is claimed to stay bounded away from zero uniformly in m. Also, the construction still assumes $1 - g = \Theta(m^{-2})$; no optimality claim about that scaling is made.*

Remark 14 (Four-phase high-load witness)**.** *At $m = 4$, the parameters $b = \kappa = \zeta = 1/4$ and $\ell = 13/320$ belong to (45). The tracker's exact saturated service is $107140/125673 \simeq 0.852529979$. The excursion bound alone gives $405/476 \simeq 0.850840336$, above the arrival rate $12277/15360$ by approximately 0.051556482.*

D.3 Why a fixed persistence target is impossible in this model

Proposition 15 (Prediction-aware converse)**.** *Under the retained event order and departure-only information, every history-only controller, with unrestricted memory and unrestricted action interface, satisfies for $t \geq 1$*

$$\Pr(A_t = S_t) \leq \max\{p, 1 - p\}, \qquad \mathbb{E}\widetilde{J}_t \leq b + (g - b)\max\{p, 1 - p\}. \tag{53}$$

This includes a relaxation granting the controller the previous hidden state S_{t-1}, but not the fresh innovation H_{t-1}. Consequently, exceeding the $(m-1)$-label menu ceiling with positive slack requires

$$\min\{p, 1-p\} < \frac{1}{m}. \tag{54}$$

For persistent phases $p \geq 1/2$, this reads $1-p < 1/m$.

Proof. Condition on the entire pre-innovation information, the previous hidden state, and any controller randomization used to select the next action. The primitive H_{t-1} is independent of this information: the departure at slot $t-1$ was generated before that innovation. Thus S_t has just the two possible values S_{t-1} and $S_{t-1}+1$, with probabilities p and $1-p$. No selected or randomized action matches with probability above their larger probability. Average the conditional inequality and apply the matched reward formula. To have a service rate strictly above $c_{m-1} = b + (g-b)(1-1/m)$ requires $\max(p, 1-p) > 1-1/m$, proving (54). □

Corollary 16 (Sharp maximal innovation order)**.** *For a strict adjacent-label separation above c_{m-1} on the persistent branch, an innovation probability of order larger than m^{-1} is impossible. The family in Theorem 12 attains such a separation with $1-p = \Theta(m^{-1})$. Thus the innovation order is matched, when the service parameters may scale as in (46). This does not identify the optimal constant or characterize the entire stability region.*

Proposition 17 (The adjacent-label gap cannot be uniform in size)**.** *If $c_{m-1} + \epsilon \leq \lambda \leq \mu - \epsilon$, then*

$$2\epsilon \leq \mu - c_{m-1} \leq g - c_{m-1} = \frac{g-b}{m} \leq \frac{1}{m}.$$

In particular, no positive ϵ independent of m can satisfy those two inequalities for all m in this normalized matched-action model.

The service gaps in Theorem 12 are of order m^{-1}, which matches this unavoidable adjacent-label scale. No global "scaling law" for arbitrary agents follows.

D.4 Quantitative robustness beyond symmetric parameters

A continuity argument alone gives a neighborhood, but not its size. The next certificate also permits hidden kernels and service matrices that break the matched-cycle symmetry.

Fix a base instance with autonomous hidden kernel H_0, hidden invariant law ν_0, service matrix R_0, and arrival rate λ_0. Its fixed identity-action tracker has saturated full phase kernel P_0. Assume

$$d_- = \lambda_0 - \Gamma_{m-1}(\nu_0, R_0) > 0, \qquad d_+ = \mu_0 - \lambda_0 > 0.$$

Choose a finite Poisson solution h_0 for which

$$d_0 + P_0 h_0 - h_0 = -d_+ \mathbf{1}, \quad d_0(s,z) = \lambda_0 - R_0(s,z), \quad M_0 = \operatorname{span}(h_0).$$

Define the hidden fundamental matrix and its zero-mass operator coefficient

$$F_0 = (I - H_0 + \mathbf{1}\nu_0)^{-1}, \qquad A_0 = \frac{1}{2}\max_{i,j}\sum_k |F_0(i,k) - F_0(j,k)|,$$

and $\omega_0 = \max R_0 - \min R_0$.

Proposition 18 (A computable asymmetric perturbation certificate)**.** *Let a perturbed product-primitive model have H', R', λ' with*

$$\max_s \mathrm{TV}(H'(s,\cdot), H_0(s,\cdot)) \leq \varepsilon_H, \quad \|R' - R_0\|_\infty \leq \varepsilon_R, \quad |\lambda' - \lambda_0| \leq \varepsilon_\lambda.$$

Suppose the hidden hold and forward-cycle probabilities stay positive, all service probabilities remain in $(0,1)$, and $\lambda' \in (0,1)$. If

$$\varepsilon_\lambda + \varepsilon_R + \omega_0 A_0 \varepsilon_H \leq d_-/2, \tag{55}$$

$$\varepsilon_\lambda + (1+M_0)\varepsilon_R + M_0\varepsilon_H \leq d_+/2, \tag{56}$$

then all $K < m$ interfaces remain excluded and the same tracker is positive recurrent with exponential queue moments from finite deterministic starts. The two respective slack bounds are $d_-/2$ and $d_+/2$.

Proof. Let ν' be the hidden invariant law. Stationarity gives the elementary perturbation identity

$$\nu' - \nu_0 = \nu'(H' - H_0)F_0.$$

The row vector before multiplication by F_0 has mass zero and TV norm at most ε_H. Decomposing its positive and negative parts shows

$$\mathrm{TV}(\nu', \nu_0) \le A_0 \varepsilon_H.$$

This is a standard finite-chain perturbation calculation; see also Schweitzer (1968). Changing the service matrix changes a fixed menu value by at most ε_R, and the oscillation of its original payoff is at most ω_0. Hence

$$\Gamma_{m-1}(\nu', R') \le \Gamma_{m-1}(\nu_0, R_0) + \varepsilon_R + \omega_0 A_0 \varepsilon_H.$$

Equation (55) gives the converse slack, and the original transient-TV and truncated-flow arguments apply.

For the tracker, couple the Bernoulli service event and the hidden transition independently. The saturated row TV distance is at most $\varepsilon_R + \varepsilon_H$. The ordinary queue drift changes by at most $\varepsilon_\lambda + \varepsilon_R$. Therefore the drift of the unchanged function $V(q, s, z) = q + h_0(s, z) - \min h_0$ at positive levels is at most

$$-d_+ + \varepsilon_\lambda + \varepsilon_R + M_0(\varepsilon_R + \varepsilon_H) \le -d_+/2.$$

The positive hold and forward-cycle probabilities retain the communication proof. Bounded increments $|\Delta V| \le 1 + M_0$ and the same exponential-transform argument as above finish the proof. □

An exact radius at the original $m = 4$ point. Here

$$d_- = 39/5120, \quad d_+ = 11420331/214481920,$$
$$\omega_0 = 13/18, \quad A_0 = 16, \quad M_0 = 320632/125673.$$

Both inequalities hold for the simultaneous choice

$$\varepsilon_H = \varepsilon_R = \varepsilon_\lambda = 1/4000.$$

Thus any admissible row-stochastic hidden kernel within this row-TV radius, any service matrix within this entrywise radius, and any arrival rate within this radius preserve the four-label minimum under departure-only feedback in this high-load regime. The kernel need not remain doubly stochastic and the service matrix need not remain matched or symmetric. This is a sufficient radius, not a maximal neighborhood.

D.5 An ML-facing extension: approximate action-law templates

Proposition 19 (Approximate screening)**.** *Fix t and let ξ_t be the actual joint law of (S_t, Z_t), with at most K labels. Suppose decoder templates $q_t(\cdot \mid z)$ satisfy*

$$\varepsilon_t = \sum_{s,z} \xi_t(s, z)\, \mathrm{TV}\left(\mathcal{L}(A_t \mid S_t = s, Z_t = z), q_t(\cdot \mid z)\right).$$

Under the manuscript's conditional service means,

$$\mathbb{E}\widetilde{J}_t \le \Gamma_K(\nu_t, R) + (r_{\max} - r_{\min})\varepsilon_t. \tag{57}$$

Consequently, for finite-mean initial queues,

$$\mathbb{E}Q_T \ge \mathbb{E}Q_0 + \sum_{t<T} \left[\lambda - \Gamma_K(\nu_t, R) - (r_{\max} - r_{\min})\varepsilon_t\right].$$

Proof. Replace the conditional action law in the expectation of $R(S_t, A_t)$ by $q_t(\cdot \mid z)$ while preserving the actual (S_t, Z_t) joint law. The change is at most the oscillation of R times ε_t. The replacement law obeys screening, so the interface-only menu proof applies to it. Telescope the queue recursion for the last claim. □

Corollary 20 (Approximate action-law covering)**.** *Let $\mathcal{G}_t$ contain the full information determining a history-based conditional action law $\pi_t(\cdot \mid \mathcal{G}_t)$, with fresh action sampling independent of S_t conditional on $\mathcal{G}_t$, so $\mathcal{L}(A_t \mid \mathcal{G}_t, S_t) = \pi_t(\cdot \mid \mathcal{G}_t)$. If there are K fixed action-distribution templates and a $\mathcal{G}_t$-measurable template assignment with expected TV error at most ε_t, then (57) holds. Thus a literal discrete bottleneck is sufficient but not necessary for a template-based upper bound.*

Proof. Use the assigned template index as an auxiliary label. Conditioning and convexity of TV show that its (S_t, Z_t)-averaged discrepancy is at most the original expected template error, then apply Proposition 19. □

Architecture interpretation. In the main construction's product-primitive model, a learned recurrent or attention encoder using only the permitted departure history, followed by a categorical label and a history-blind decoder with fresh action sampling, satisfies the joint theorem. A *probability mixture* of K history-blind action distributions also satisfies it under the same information conditions, even when the gate is continuous, because it admits a latent categorical-label factorization. Averaging logits is not in general the same as mixing action probabilities. History-dependent experts, action-visible recurrent decoder state, observation-dependent action masks, and continuous action arguments require separate accounting. A finite decoder side state with L values gives the coarse bound for at most KL effective labels, when no other input bypasses that pair. An empirical fit of K templates on finitely many histories alone is not a distribution-free covering certificate.

At the original $m = 4$, $K = 3$ point, the under-capacity obstruction survives any uniform TV error strictly below

$$\frac{\lambda - c_3}{g - b} = \frac{27}{2560} \simeq 0.010546875.$$

This is a converse robustness threshold, not an achievability claim at or above that amount of leakage.

D.6 Finite verification and phase certificates

The supplementary `evidence/theory/robust/` contains the verifier and stored output for 520 exact rational saturated chains and 1,040 parameter configurations at the two load endpoints, for $m = 2, \ldots, 64, 128, 256$. All corner inequalities and four exact Poisson certificates pass. A separately assembled float64 linear solve checks 248 chains. These finite checks support the implementation of the formulas; the proofs above establish the continuum and all-size statements. No new queue simulation or learning run is represented by these certificates.

E A joint innovation–interface service certificate

This section proves Theorem 2 and applies it to histories that do not reveal fresh hidden innovations. The product primitives and event order of the queue construction imply its conditional laws, but the theorem itself needs only those laws and the conditional service means. Write H for the hidden transition matrix, to distinguish it from the Bernoulli increment H_t.

Proof of Theorem 2. Let $\xi(s, z) = \Pr(S_{t-1} = s, Z_t = z)$. First average service over (S_t, A_t). Factoring the joint law of this pair using the two conditional laws (9)–(10) then gives

$$\mathbb{E}\widetilde{J}_t = \sum_{s,z} \xi(s, z) \sum_a q_t(a \mid z)(\mathsf{H}R)(s, a).$$

For each non-null label choose an action maximizing its contribution. The same linearity argument as in Theorem 1, now with service matrix $\mathsf{H}R$, gives the first inequality in (11). For a fixed action menu M, the elementary inequality

$$\max_{a \in M} \sum_{s'} \mathsf{H}(s, s') R(s', a) \le \sum_{s'} \mathsf{H}(s, s') \max_{a \in M} R(s', a)$$

remains true after averaging over ν and maximizing over menus. This proves the interface-only comparison in (11). Maximizing over all actions instead of a menu gives its prediction-only comparison. The argument allows arbitrary correlation between S_{t-1} and Z_t. In the retained queue

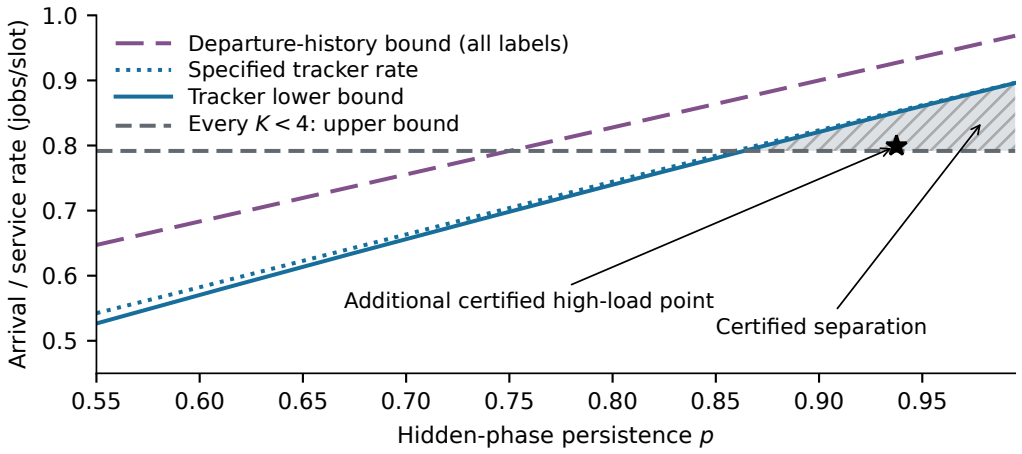

Figure 4: Analytic certificates at $m = 4$, $b = 1/4$, $g = 35/36$, varying persistence. The shaded interval lies strictly between the smaller-interface ceiling and the excursion lower bound on tracker service. It is a certified separation region, not the full stability region. The departure-history upper bound applies to all history-only controllers; the dotted stationary service concerns the specified tracker. The star marks an additional certified high-load point at $p = 15/16$ and $\lambda = 12277/15360$, distinct from the main intervention load 0.78. Unshaded gaps between sufficient and necessary bounds remain unclassified by these bounds.

model, $Y_{t-1} = D_{t-1}$ is generated before H_{t-1}. Conditional on the previous hidden state, this independent increment is not revealed by the history used to select Z_t, so (9) holds. An encoder additionally granted S_{t-1} still satisfies it. An encoder granted the *current* state S_t need not; its valid reference remains the interface-only menu bound. Autonomy of a hidden marginal alone, without the conditional independence above, is not a replacement for this hypothesis.

Finite-time queue consequence and initialization. The joint bound applies from slot one. Apply Theorem 1 at slot zero, then telescope $Q_{t+1} - Q_t = X_t - D_t$ with $D_t \le \widetilde{J}_t$ to obtain (12). In particular, when $\nu_0 = \pi$ is invariant, the right side equals $\mathbb{E}Q_0 + \lambda - \Gamma_K(\pi, R) + (T-1)[\lambda - \Psi_K(\pi, \mathsf{H}, R)]$. For a nonstationary initial law and a finite irreducible aperiodic hidden chain, the same TV argument as in Theorem 1 bounds the transient correction by the oscillation of $\mathsf{H}R$ times $\sum_{s \ge 0} \mathrm{TV}(\nu_s, \pi) < \infty$, together with the initial-slot term. Thus a strict gap below demand yields a linear expected-backlog lower bound up to a finite transient constant.

For a homogeneous compliant augmented realization with unique hidden invariant law π, $\lambda > \Psi_K(\pi, \mathsf{H}, R)$ excludes an invariant probability law. Under a hypothetical stationary realization, apply Theorem 2 to two successive slots. Truncated queue balance gives $\lambda = \mathbb{E}D_t \le \mathbb{E}\widetilde{J}_t \le \Psi_K$, a contradiction, without assuming a finite stationary queue mean. This uses the same conditional laws under that stationary realization.

Corollary 21 (Information age and menu capacity)**.** *Let $\pi\mathsf{H} = \pi$, $K \ge 1$, and $C_K^{(d)} = \Gamma_K(\pi, \mathsf{H}^d R)$ for integers $d \ge 0$. Write $c_{\mathrm{blind}} = \max_a \sum_s \pi(s) R(s, a)$. Then*

$$c_{\mathrm{blind}} = C_1^{(d)} \le C_K^{(d+1)} \le C_K^{(d)}.$$

If H is irreducible and aperiodic, then, uniformly in K,

$$0 \le C_K^{(d)} - c_{\mathrm{blind}} \le (r_{\max} - r_{\min}) \max_s \mathrm{TV}(\mathsf{H}^d(s, \cdot), \pi) \longrightarrow 0.$$

For $t \ge d$, suppose the controller satisfies, at feasible values,

$$\Pr(S_t = s' \mid S_{t-d} = s, Z_t = z) = \mathsf{H}^d(s, s'), \qquad \Pr(A_t = a \mid S_{t-d}, S_t, Z_t = z) = q_t(a \mid z).$$

Then $\mathbb{E}\widetilde{J}_t \le \Gamma_K(\nu_{t-d}, \mathsf{H}^d R)$.

Proof. Stationarity gives $C_1^{(d)} = c_{\mathrm{blind}}$, and every menu budget admits a singleton. For each menu M,

$$\max_{a\in M}(\mathsf{H}^{d+1}R)(s,a) \le \sum_u \mathsf{H}(s,u)\max_{a\in M}(\mathsf{H}^dR)(u,a).$$

Average under π and maximize over M to obtain monotonicity. The total-variation inequality bounds each entry of H^dR within the stated error of $\sum_s \pi(s)R(s,a)$; menu maximization preserves this bound. Finite-chain mixing proves convergence. For the service bound, expand the marginal law of (S_t, A_t) through (S_{t-d}, Z_t) and apply the menu argument to H^dR. □

The cases $d = 0$ and $d = 1$ recover the interface-only and MINT ceilings. Each fixed-budget ceiling is nonincreasing with information age. If H is irreducible and aperiodic and $\lambda > c_{\mathrm{blind}}$, there is a finite d_0 such that $C_K^{(d)} < \lambda$ for every $d \ge d_0$ and every K. For any such fixed delay satisfying the displayed conditional laws, stationary hidden initialization and $\mathbb{E}Q_0 < \infty$ give linear expected-backlog growth, with a finite correction for the first d slots. Truncated stationary flow balance excludes invariant laws for homogeneous augmented realizations with a fixed label-only decoder satisfying these laws under stationarity. Ordinary departure feedback between $t-d$ and t can reveal intervening innovations, so the delayed-information laws cannot be inferred from hidden-chain autonomy alone.

Corollary 22 (Closed form for a persistent matched cycle)**.** *Let $p \ge 1/2$, $k = \min(K, m)$, and retain the matched service matrix. For uniform hidden phase and the hold/forward kernel H_p,*

$$\Psi_K(u_m, \mathsf{H}_p, R) = b + \frac{g-b}{m}\big[pk + (1-p)\min(k, m-k)\big]. \tag{58}$$

For $0 \le \alpha \le 1$ and $\mathsf{H}_\alpha = \alpha\mathsf{H}_p + (1-\alpha)\mathbf{1}u_m^\top$,

$$\Psi_K(u_m, \mathsf{H}_\alpha, R) = b + (g-b)\left[\frac{1-\alpha}{m} + \frac{\alpha}{m}\big[pk + (1-p)\min(k, m-k)\big]\right]. \tag{59}$$

Proof. Fix a menu of k distinct actions; allowing unused extra labels loses nothing because adding actions cannot reduce a menu value. In the cycle, the part of $(\mathsf{H}_pR)(s,a)$ above b equals $(g-b)p$ for $a = s$ and $(g-b)(1-p)$ for $a = s+1$. Since $p \ge 1/2$, each phase in the menu contributes p. A phase outside the menu contributes $1-p$ precisely when its successor is in the menu. The number of these outside-to-inside edges is at most both k and $m-k$. This maximum is attained by isolating the minority set of positions on the cycle; at $k = m$ it is zero. Averaging proves (58), including $m = 2$ and $p = 1/2$. The iid component contributes $(g-b)/m$ independently of the selected action; it therefore commutes with the menu maximum, proving (59). □

For the pure cycle, the tightening has the explicit size

$$\Gamma_K(u_m, R) - \Psi_K(u_m, \mathsf{H}_p, R) = \frac{(g-b)(1-p)}{m}(2k-m)_+.$$

It need not be strict for small menus. At $p = 1/2$, the previous-state-oracle ceiling saturates once $k \ge \lceil m/2 \rceil$: additional labels cannot reveal which of the two equally likely successor phases occurred. This is a statement about the certificate and its previous-state oracle, not general achievability from departures.

Quantitative gain at a higher load. For $m = 4$, $p = 15/16$, $b = 1/4$, $g = 35/36$ and $\lambda = 12277/15360$,

$$\Gamma_3 = \frac{19}{24}, \quad \Psi_3 = \frac{443}{576}, \quad \lambda - \Psi_3 = \frac{1391}{46080}.$$

The MINT service shortfall is $107/27 \simeq 3.963$ times the interface-only shortfall $\lambda - \Gamma_3 = 39/5120$. Its ceiling Ψ_3 is strictly smaller than both the three-label interface ceiling and the unrestricted prediction ceiling. At $K = m$ it recovers the persistent-branch prediction ceiling $b + (g-b)p$. At $\alpha = 0$ it gives $b + (g-b)/m$ for every K, incorporating the iid control into the same certificate. For $K = 4$, exceeding this demand requires $\alpha > 1307/1760 \simeq 0.74261$; this is a necessary condition, not a claim of history-observable achievability.

Corollary 23 (Strict gap between menu clearance and history feasibility)**.** *Keep $m = 4$, $b = 1/4$, $g = 35/36$ and $p = 15/16$, but set $\lambda^\dagger = 39/50$. The minimum strictly capacity-clearing menu has three labels, yet the minimum label-only interface admitting a positive-recurrent departure-history controller has four labels.*

Proof. The exact inequalities are

$$\Gamma_2 < \lambda^\dagger < \Gamma_3 = 19/24, \qquad \Psi_3 = 443/576 < \lambda^\dagger, \qquad \lambda^\dagger - \Psi_3 = 157/14400.$$

Thus the interface-only menu criterion clears at three labels, as does the unrestricted prediction ceiling $89/96$, but the joint certificate excludes invariant laws for every compliant history-based interface of at most three labels. The specified four-label tracker's saturated service is independent of arrivals and exceeds $12277/15360$, hence also $39/50$. The same communicating boundary and Poisson-corrected Foster argument applies because $0 < 39/50 < 1$. It therefore yields a positive-recurrent four-label controller at this load. □

An exact witness makes the constructive claim checkable without simulation. For difference phase $d = (s - a) \bmod 4$, take

$$h = (0,\ 36712/41891,\ 220064/125673,\ 320632/125673), \qquad V(q, s, a) = q + h_d.$$

The stationary service is $107140/125673$, and the drift of V at every positive queue level and each of the 16 full phases equals

$$39/50 - 107140/125673 = -455753/6283650 < 0.$$

All primitive probabilities lie strictly between zero and one, so the finite boundary and communicating-class argument used above applies. This rational certificate accompanies the argument for all histories; it is not an extra simulated trajectory.

This analytic load example uses exact model calculations. It shows that the joint certificate can reject a controller class that both separate tests fail to exclude, rather than merely strengthening an already negative service margin.

Role of the certificate. The MINT ceiling is attained by an oracle that sees S_{t-1} and chooses an optimizing fixed menu; the departure-only controller may not realize that oracle. The gap between Ψ_K and a learned controller consequently contains both any residual observation constraint and optimization error. It is not identified as a pure optimization gap. The current-phase oracle can exceed Ψ_K without violating it: it does not satisfy (9).

The verifier and stored output in `evidence/theory/innovation/` cover 396 exact matched-cycle cases through $m = 9$, 120 rational randomized kernels, and all 70,980 deterministic encoder/decoder pairs at $m = 4$. These checks accompany, rather than replace, the arbitrary-size argument above.

Finite decoder side information. If the decoder additionally reads $W_t \in [L]$ through fixed rows $q_t(a \mid z, w)$, the combined label (Z_t, W_t) has at most KL values. Applying the same screening proof gives the interface ceiling Γ_{KL}. The joint ceiling Ψ_{KL} also follows if both conditional-independence conditions above hold with this combined label; current-phase side information need not satisfy them. This counting reduction does not establish that one or two memory bits suffice for stabilization. An externally prescribed clock is already covered by time-dependent rows in the finite-horizon bound; history-correlated decoder memory changes the information contract.

F Certified intervention intervals and closed-loop evaluation

We prove Theorem 4 and report the corresponding frozen-controller evaluation. This extends the single-load certificate in Appendix E. The class-wide bound excludes undersized interfaces and the recurrence proofs establish feasibility for the explicit interventions; the finite queue runs show their workload.

F.1 ARBITRARY-SIZE INTERVENTION INTERVALS

For $m \geq 3$, take the physical parameters in (15) and choose $\lambda \in \mathcal{I}_m = (\Psi_{m-1}, c_{m-1})$. The matched-cycle closed form yields

$$\begin{aligned}
\Psi_{m-1} &= b + \frac{g-b}{m}\big[m - 1 - (1-p)(m-2)\big], \\
c_{m-1} - \Psi_{m-1} &= (g-b)\kappa\frac{m-2}{m^2} > 0, \\
\Psi_{m-1} - c_{m-2} &= \frac{g-b}{m}\big[1 - \kappa(m-2)/m\big] > 0, \\
b + (g-b)p - c_{m-1} &= \frac{(g-b)(1-\kappa)}{m} > 0.
\end{aligned}$$

In particular $c_{m-2} < \Psi_{m-1} < \lambda < c_{m-1} < b + (g-b)p$. At $K = m-1$, both separate ceilings clear demand, while the joint certificate excludes every $K \leq m-1$ departure-history controller. Because $g - b \geq 1/2$, $\kappa \geq 1/8$ and $(m-2)/m \geq 1/3$,

$$\frac{1}{48m} \leq |\mathcal{I}_m| \leq \frac{1}{4m}, \qquad m \geq 3.$$

The interval therefore has the same inverse-size order as the adjacent-label capacity gap; this does not assert a size-independent physical volume. For $m = 2$, the displayed interval degenerates and no strict interval claim is made.

The saturated tracker rate does not depend on arrivals. Proposition 3 gives $\mu > c_{m-1} > \lambda$ and, with its Poisson correction and finite communicating boundary, a positive-recurrent m-label tracker at these lower loads. Thus exactly m labels are required under departure history.

For the observation intervention, let the encoder read current S_t and use $m-1$ deterministic action rows $0, \ldots, m-2$. Match those states, and select action zero in state $m-1$. Its service vector is $(g, \ldots, g, b)$ with uniform stationary mean c_{m-1}. The explicit correction

$$h_{\mathrm{obs}}(s) = \frac{s(g-b)}{m(1-p)} = \frac{s(g-b)}{\kappa}, \quad s = 0, \ldots, m-1,$$

satisfies $(I - \mathsf{H}_p)h_{\mathrm{obs}} = c_{m-1}\mathbf{1} - r$: for $s < m-1$ the difference is $-(g-b)/m$, and for $s = m-1$ it is $(m-1)(g-b)/m$. Hence $q + h_{\mathrm{obs}}(s)$ has negative drift $\lambda - c_{m-1}$ at every positive level. Positive hold/forward probabilities and strictly interior service/arrival probabilities connect the finite boundary and every positive level. The Foster criterion proves positive recurrence. The original oracle-valid menu bound excludes all $K \leq m-2$ since $c_{m-2} < \lambda$. Thus exactly $m-1$ labels are necessary with current-phase observation. This proves Theorem 4. The change in observation is explicit; it is not an uncounted path into the decoder.

F.2 A FOUR-PHASE WORKED WITNESS AND EXACT QUEUE CONSEQUENCE

Fix $m = 4$, $b = 1/4$, $g = 35/36$, $p = 15/16$ and the specified product primitives. Let

$$\mathcal{I} = \left(\frac{443}{576}, \frac{19}{24}\right), \qquad |\mathcal{I}| = \frac{13}{576}.$$

The endpoints are excluded; no critical-load classification is needed. For every $\lambda \in \mathcal{I}$,

$$\Gamma_2 = \frac{11}{18} < \Psi_3 = \frac{443}{576} < \lambda < \Gamma_3 = \frac{19}{24} < \Psi_4 = \frac{89}{96}. \tag{60}$$

Thus both separate tests clear the load while the joint test excludes it.

Unchanged observations and at most three labels. Departure history leaves the current innovation independent of the encoder given the previous phase, as required by Appendix E, even with arbitrary internal memory and nonstationary processing. For Q_0 of finite mean and uniform S_0, applying the interface-only menu bound to slot zero and the joint bound thereafter gives

$$\mathbb{E}Q_T \geq \mathbb{E}Q_0 + T\left(\lambda - \frac{443}{576}\right) - \frac{13}{576}, \qquad T \geq 1. \tag{61}$$

Monotonicity strengthens the exclusion for smaller K; finite-chain mixing adds a finite correction for nonuniform hidden initialization. Under a hypothetical invariant law of a compliant homogeneous augmented realization, the hidden marginal is uniform. Applying stationarity to $\min(Q, N)$ and dominated convergence to bounded increments gives $\mathbb{E}D_t = \lambda$, contradicting $\mathbb{E}D_t \le \Psi_3 < \lambda$ without assuming a finite invariant queue mean. At $\lambda = 39/50$, the coefficient in (61) is $157/14400$; at $T = 100{,}000$, the lower bound above $\mathbb{E}Q_0$ is $209329/192 \simeq 1090.2552$.

Interface intervention: four labels, original feedback. Keep the departure-driven four-state recurrence and map each label to the physical action with the same index. Define $d = (s - c) \bmod 4$ and

$$\mu_{\mathrm{tr}} = \frac{107140}{125673}, \qquad h = \left(0, \frac{36712}{41891}, \frac{220064}{125673}, \frac{320632}{125673}\right), \qquad V(q, s, c) = q + h_d.$$

The saturated difference-phase calculation gives $(I - P)h = \mu_{\mathrm{tr}} - \theta$, where $\theta_0 = g$ and $\theta_1 = \theta_2 = \theta_3 = b$. At every $q \ge 1$, the actual recurrence coincides with the saturated one, so

$$\mathbb{E}[\Delta V \mid q, s, c] = \lambda - \mu_{\mathrm{tr}} < 0,$$

because $\mu_{\mathrm{tr}} > 19/24 > \lambda$. Strictly interior service, arrival, hold and forward probabilities preserve the communication proof of Appendix C. Its finite boundary and finite sublevel sets of V give recurrence; bounded increments give exponential queue moments for each fixed interior load by the same exponential-transform argument. This intervention preserves feedback and the $M = 4$ internal update; it changes the four-memory compressed output map from $(0, 1, 2, 2)$ to $(0, 1, 2, 3)$.

Observation intervention: three labels, current-phase access. Grant the encoder S_t before label selection. Use three deterministic decoder rows, for actions $0, 1, 2$, and encode states $0, 1, 2, 3$ to actions $(0, 1, 2, 0)$. Its service vector is $r = (g, g, g, b)$ and its uniform stationary service is

$$\mu_{\mathrm{obs}} = \frac{3g + b}{4} = \frac{19}{24}.$$

For the hold/forward hidden kernel, take

$$h_{\mathrm{obs}} = \left(0, \frac{26}{9}, \frac{52}{9}, \frac{26}{3}\right).$$

Direct substitution yields $(I - \mathsf{H}_p)h_{\mathrm{obs}} = \mu_{\mathrm{obs}}\mathbf{1} - r$. Thus $V_{\mathrm{obs}}(q, s) = q + h_{\mathrm{obs}}(s)$ has drift $\lambda - 19/24 < 0$ at every positive queue state. Service with arrival and a hidden forward step changes s at fixed positive q; positive-probability queue increases and decreases connect these phases to the finite boundary. The same Foster argument proves recurrence. Current-phase access violates the unobserved-innovation condition (9) but obeys the original three-label action screening. With only two labels the original oracle-valid bound $\Gamma_2 = 11/18 < \lambda$ rules out stability even under current-phase access. Consequently, the minimum stable label counts are exactly four under departure history and three under the changed observation contract.

An independent-phase observation obstruction. Interpolate between the persistent kernel and independent uniform phases: $\mathsf{H}_\alpha = \alpha\mathsf{H}_p + (1 - \alpha)\mathbf{1}u_4^\top$, $0 \le \alpha \le 1$. The three-label interface-only bound is unchanged, while $\Psi_4 = b + (g - b)[(1 - \alpha)/4 + \alpha p]$. At $\alpha = 0$, the kernel is $\mathbf{1}u_4^\top$; retain uniform hidden initialization. Each phase is freshly sampled independently, so the current phase is independent of all past public information. Every history-only controller therefore has potential service $b + (g - b)/4 = 31/72$, regardless of its label budget. All loads in $\mathcal{I}$ exceed this value, so enlarging labels alone cannot restore stability. The current-phase three-label policy still has service $19/24$. The correction $h_{\mathrm{iid}} = (0, 0, 0, 13/18)$ gives $(I - \mathbf{1}u_4^\top)h_{\mathrm{iid}} = (19/24)\mathbf{1} - r$, proving recurrence through the same finite boundary even without temporal predictability.

What the interval decides. Within this region, the explicit four-label departure tracker and three-label current-phase policy each stabilize the queue. The joint class bound excludes stabilization by any memory or optimization change that retains $K = 3$ and departure-history feedback. The independent-phase condition shows that label enlargement itself can be ineffective. Clearance elsewhere remains undecided without a constructive policy argument.

F.3 PROTOCOL FROZEN BEFORE THE NEW QUEUE SIMULATIONS

The standalone evaluator imports no project simulation or learning classes. The model, four loads, controller tables, 32 seeds 261301–261332, 100,000-slot horizon and 1,000-slot trace stride were serialized with a source hash before these runs. The interval and controllers were already known; this protocol does not predate that research.

The four controller tables are the previous selected three-state FSC, four-state/three-label example, identity four-label tracker and current-phase three-menu reference. Their prior selection and initialization are frozen across the load grid, with no selection from new results. The three-state map came from a historical exhaustive search over controllers with at most three reachable states, using a 10,000-slot phase-zero objective; the four-state compressed map came from a separate search over 232 output maps with its update fixed. Neither selection targeted this four-load grid, and the two examples do not form a controlled memory-only comparison. The regular recurrence is $C_{t+1} = C_t + 1 - D_t \pmod M$, including when the queue is empty. Actions use the current recurrence state, or S_t for the current-phase observation intervention. Independent service, arrival and hidden-transition uniform streams are drawn from separate PCG64 children of each seed. The same primitive bank is reused across controllers and loads within a seed. At iid hidden transitions that same uniform stream is instead mapped to a uniform phase; no inference of independence between the two kernel conditions is made.

Each trajectory begins with $Q_0 = C_0 = 0$ and an independently uniform hidden phase. Potential service is sampled each slot, actual departure is zero at the empty boundary, and queues are never truncated or reset. Every endpoint obeys $Q_T = \sum_t X_t - \sum_t D_t$ exactly. We report terminal, time-average and maximum queue, actual departure and arrival rates, and $(Q_T - Q_{T/2})/(T/2)$. None is relabeled as a statistical recurrence test. Saturated service is solved separately on the finite (S, C) chain; it is not estimated by conditioning a reflected-queue trajectory on being nonempty.

There are four loads times four methods times 32 seeds, plus two iid methods at load 0.78 times 32 seeds: 576 trajectories and 57.6 million queue transitions. Two-sided t intervals are computed per cell across 32 independent environment seeds; methods and loads sharing a seed are not independent replications. Trajectory bands are central 80% seed ranges, not confidence intervals for every time point. The archive includes per-trajectory statistics, endpoints, frozen selection references and queue traces sampled every 1,000 slots for all 576 trajectories; the simulator and seeds permit full replay. No neural training occurs in this study.

F.4 ALL LOAD CELLS, INCLUDING THE NEAR-BOUNDARY ORACLE

Table 3 retains all 18 cells; Figure 5 shows all four persistent-load points.

F.5 INDEPENDENT ARITHMETIC AND IMPLEMENTATION CHECKS

The verifier directly enumerates service outcomes and hidden innovations for all 16 full tracker phases, at each of the four loads and both interval endpoints. All 96 tracker drift evaluations equal $\lambda - \mu_{\mathrm{tr}}$. It separately checks the four oracle phase drifts at these six loads; the upper interval endpoint has zero oracle drift and is not claimed stable here. An independent exact-rational stationary solve evaluates all four controller kernels under both persistent and iid hidden dynamics. The two history-only three-label rates under persistence are respectively $113299973873237/157967273215160$ and $1618079952140895/2279293250895866$; the tracker and oracle rates are $107140/125673$ and $19/24$.

Independent replay checks the reported queue statistics and empty-boundary identities against the stored trajectories. These finite checks validate the implementation; the continuum and recurrence statements follow from the proofs above.

The arbitrary-size conclusion follows from the inequalities above; the four-mode certificates and trajectory replays do not replace that proof.

Table 3: All intervention cells. Q_T is a finite-horizon mean over 32 seeds; the halfwidth is a two-sided 95% t interval. The queue estimates measure fixed-policy workload; positive recurrence for the two interventions is proved separately.

Load / kernel	Controller	Saturated rate	$\mathbb{E}Q_T$ estimate	95% halfwidth
0.775	Selected FSC ($M = K = 3$)	0.717237	5724.09	154.97
0.775	4-state example / departures	0.709904	6449.09	160.10
0.775	Four-label tracker	0.852530	4.72	1.72
0.775	Current-state / three-label	0.791667	41.69	17.09
0.780	Selected FSC ($M = K = 3$)	0.717237	6221.91	154.20
0.780	4-state example / departures	0.709904	6947.81	159.51
0.780	Four-label tracker	0.852530	5.22	1.66
0.780	Current-state / three-label	0.791667	58.84	22.16
0.785	Selected FSC ($M = K = 3$)	0.717237	6722.75	154.84
0.785	4-state example / departures	0.709904	7449.09	160.01
0.785	Four-label tracker	0.852530	6.00	1.97
0.785	Current-state / three-label	0.791667	104.25	34.90
0.790	Selected FSC ($M = K = 3$)	0.717237	7230.97	153.24
0.790	4-state example / departures	0.709904	7957.50	158.40
0.790	Four-label tracker	0.852530	5.84	1.93
0.790	Current-state / three-label	0.791667	230.88	69.08
0.780 / iid	Four-label tracker	0.430556	34919.38	72.64
0.780 / iid	Current-state / three-label	0.791667	11.97	5.17

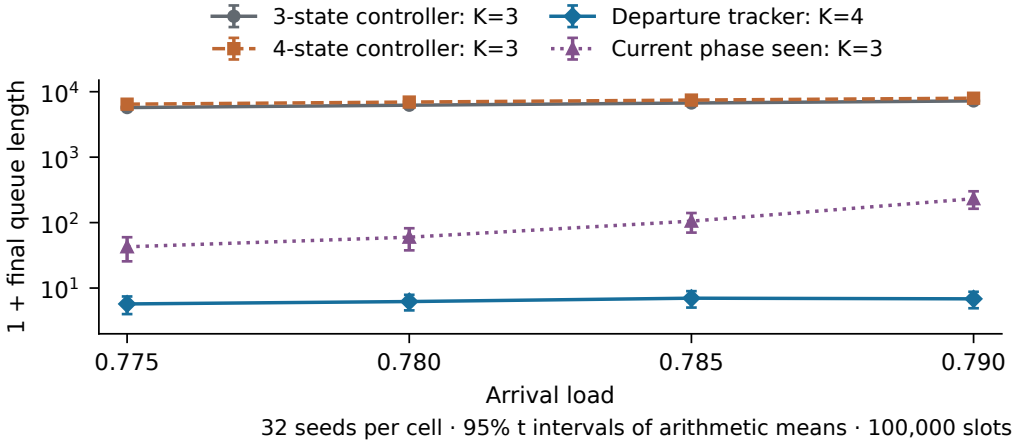

Figure 5: All four fixed persistent-load cells per method. Points are $1+$ the arithmetic mean final queue after 100,000 slots; bars are two-sided 95% t intervals over 32 environment seeds, transformed with the same $+1$ offset for the log display. Lines only guide the eye. The current-phase three-label policy has larger finite queues near its service boundary than the four-label departure tracker. Both are positive recurrent; restored stability does not imply equal delay.

G THE JOINT SEPARATION BEYOND THE CYCLE

The main construction uses a hold-or-forward hidden cycle. Here we give a finite-state criterion, evaluable by menu enumeration and stationary linear solves, and certify the same intervention distinction in two previously reported nearby environments. Their hidden kernel allows a jump to *every* phase, so neither instance is covered by the hold-or-forward closed form. The second also has unequal matched-action service rates.

Proposition 24 (Finite-state joint-intervention criterion)**.** *For the no-bypass K-label interface of Section 2, let $m \geq 3$, $|\mathcal{S}| = |\mathcal{A}| = m$, and let the autonomous hidden kernel H have strictly positive entries and invariant law ν. Suppose $R(s,a)$ and λ are strictly between zero and one. Conditional on the full predecision past and (S_t, A_t), let arrival, potential service and the next hidden transition be fresh independent draws with the stated parameters, in the order action, potential*

service, actual departure, arrival, then hidden transition. Use the departure-history tracker $A_t = Z_t$, $Z_{t+1} = Z_t + 1 - D_t \pmod m$, *and let* μ_{tr} *be its saturated potential-service rate, obtained from the invariant law of its finite positive-level phase* (S_t, Z_t). *If*

$$\max\{\Psi_{m-1}(\nu, \mathsf{H}, R), \Gamma_{m-2}(\nu, R)\} < \lambda < \min\{\Gamma_{m-1}(\nu, R), \mu_{\mathrm{tr}}\}, \tag{62}$$

then the minimum stable label count is m *under departure-only feedback and* $m-1$ *when the encoder is instead granted the current hidden state. Both separate* $m-1$*-label interface-only and unrestricted prediction-only upper bounds exceed* λ*, whereas the joint bound excludes all departure-history controllers with at most* $m-1$ *labels. The minimum counts concern positive-recurrent homogeneous controllers under their respective observation contracts; the exclusions also give linear expected-backlog growth up to a finite initialization correction.*

Proof. Since every entry of H is positive, the hidden chain mixes geometrically. The joint menu theorem applies to departure-only histories: the fresh hidden transition is independent of the label conditional on S_{t-1}. Its bound, the left side of (62), and monotonicity in K exclude every $K \le m-1$ history controller. The finite-time correction for an arbitrary initial hidden law is summable. The original menu bound excludes every $K \le m-2$ controller even if the encoder observes S_t.

At positive queue levels the tracker has phase kernel, for $i = (s, z)$ and $j = (s', z')$,

$$P_{ij} = \mathsf{H}(s, s')\big[R(s, z)\mathbf{1}\{z' = z\} + (1 - R(s, z))\mathbf{1}\{z' = z + 1 \pmod m\}\big]. \tag{63}$$

Positive entries of H and interior service probabilities make this finite chain irreducible and aperiodic. Write $\eta P = \eta$, $r(s, z) = R(s, z)$, and $\mu_{\mathrm{tr}} = \eta r$. The centered finite-chain Poisson equation has a solution $(I - P)h = \mu_{\mathrm{tr}}\mathbf{1} - r$. Therefore $V(q, s, z) = q + h(s, z) - \min h$ has conditional drift $\lambda - \mu_{\mathrm{tr}} < 0$ at every positive queue level.

For current-phase access, choose a fixed $(m-1)$-action menu attaining $\Gamma_{m-1}(\nu, R)$ and in each state select its best action. Its saturated reward vector r_{obs} has mean $\nu r_{\mathrm{obs}} = \Gamma_{m-1}$. Solving $(I - \mathsf{H})h_{\mathrm{obs}} = \Gamma_{m-1}\mathbf{1} - r_{\mathrm{obs}}$ gives negative corrected drift $\lambda - \Gamma_{m-1} < 0$ at positive queue levels. In both chains, positive arrivals and interior service probabilities connect the finite empty-queue boundary to positive levels; finite sublevel sets and the Foster criterion give positive recurrence. Finally the unrestricted prediction-only ceiling is $\Psi_m(\nu, \mathsf{H}, R)$, which is at least μ_{tr} by the joint bound applied to the saturated tracker. Hence it also exceeds λ. □

Two exact off-cycle witnesses. Take $m = 4$, $\lambda = 39/50$, and a noncyclic hidden kernel with $\mathsf{H}(s, s) = 47/50$ and $\mathsf{H}(s, s') = 1/50$ for $s' \neq s$. The invariant hidden law is uniform. In the first instance, service is $35/36$ when action and phase match and $1/4$ otherwise. In the second, the four matched rates are $(47/50, 24/25, 49/50, 99/100)$, with the same unmatched rate $1/4$. These are exactly the noncyclic and heterogeneous models already used for the descriptive Full-head neural transfer readouts in Appendix I; those readouts are not stability proofs.

Table 4: Exact off-cycle certificates at the same load. Γ_2 and Ψ_3 are class ceilings; Γ_3 is attained with current-phase access, and $\mu_{\mathrm{tr}} > 4/5$ is attained by the departure tracker. The script emits exact tracker rates and Poisson solutions.

Instance	Γ_2	Ψ_3	Γ_3	μ_{tr}
Noncyclic, matched	$11/18$	$1373/1800$	$19/24$	$> 4/5$
Noncyclic, unequal rates	$247/400$	$383/500$	$159/200$	$> 4/5$

For both rows, the table establishes $\Gamma_2 < \Psi_3 < \lambda < \Gamma_3 < 4/5 < \mu_{\mathrm{tr}}$. The prediction-only ceilings are $209/225$ and $18489/20000$, respectively, so neither separate test rules out the load. To verify the joint entries directly, write $\delta_a = R(a, a) - 1/4$. For the listed rates, an included phase prefers its matching action. Thus any three-action menu M has

$$\sum_s \nu(s) \max_{a \in M} (\mathsf{H}R)(s, a) = \frac{1}{4} + \frac{1}{4}\left(\frac{47}{50}\sum_{a \in M}\delta_a + \frac{1}{50}\max_{a \in M}\delta_a\right).$$

The three largest δ_a maximize this expression and also maximize Γ_3. For the tracker, solve the 16-state rational system $\eta P = \eta$, $\eta\mathbf{1} = 1$, using (63) and check $\eta r > 4/5$. The script additionally

verifies both Poisson-equation residuals exactly. Thus all comparisons needed by Proposition 24 are strict, without relying on finite-horizon queue trajectories.

The third nearby model reported in Appendix I has asymmetric hold-or-forward persistence. At the same load, its three-label joint ceiling is $12677/14880 > \lambda$, so this converse does not exclude three labels there. Its learning readouts remain reported; it is not counted as a third joint-separation witness.

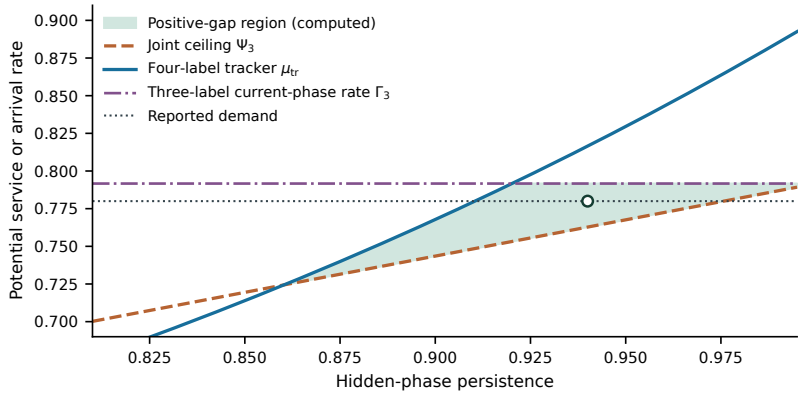

Figure 6: Computed region for the all-to-all kernel $\mathsf{H}(s,s) = p$, $\mathsf{H}(s,s') = (1-p)/3$ for $s' \neq s$. Shading visualizes $\Psi_3(p) < \lambda < \min\{\Gamma_3, \mu_{\mathrm{tr}}(p)\}$; Γ_2 lies below it. Proposition 24 certifies points satisfying these strict inequalities. Curves and shading are numerical; the circle and Table 4 are checked exactly. The region is sufficient, not exhaustive.

H Reward-only finite-controller search and post-training certification

This study connects interaction-based learning to the queue theorem without teacher actions or a fixed teacher recurrence. It learns controller tables from sampled rewards. Only after search are the frozen controllers checked against the known queue model; search does not access model parameters or exact service rates. This study does not propose a general-purpose RL algorithm. The learning protocol was frozen before any new optimizer update; the empirical queue-statistic and bootstrap plan was recorded after training and initial certificate inspection but before any confirmation queue evaluation. Neither is represented as an externally registered protocol.

H.1 Execution contract and route selection

The environment uses only $m = 4$, $b = 1/4$, $g = 35/36$, $p = 15/16$ and $\lambda = 39/50$. All three label conditions have four internal memory states. The trainable maps are

$$Z_t \sim E(\cdot \mid C_t), \qquad A_t \sim W(\cdot \mid Z_t), \qquad C_{t+1} \sim T(\cdot \mid C_t, D_t).$$

The memory update follows the actual departure; it never receives potential service at an empty queue. The decoder receives only the current label, not memory, queue length, hidden phase, history, logits or a continuous gate. There is no critic, teacher, imitation target, model-based reward gradient or pretrained recurrence. The learning implementation returns only $(C_t, Z_t, A_t, D_t, C_{t+1})$ to the score-function trainer. The population scorer returns only sampled departure averages. Independent queue and hidden-state fields are used by the simulator and by post-training audits, not as policy inputs or optimization targets.

For a horizon H, each training episode begins at $Q_0 = H + 1$, with independent uniform hidden and memory states. Therefore actual departures equal potential service throughout that finite episode. The reward is $H^{-1} \sum_t D_t$. This high-backlog training distribution estimates saturated performance; it does not train online stabilization from an empty queue. The queue still obeys the original update, with no cap, within-episode reset or privileged feedback. Empty-boundary correctness and deployment are checked separately.

The first route is a table-based REINFORCE learner (Williams, 1992): 1,500 Adam ascent updates, 128 episodes of length 128 per update, learning rate 0.03, Gaussian initial logits with standard deviation one, and no entropy bonus. All $8K + 32$ logits are optimized. Return-to-go weights encoder and action scores; a memory score after D_t receives only rewards strictly after time t. A leave-one-episode-out baseline is computed separately for each time index. An independent finite-horizon dynamic program and directional finite differences check the gradient, but neither is called during optimization. The diagnostic source and stored output are included in `evidence/fsc/code/test_contract.py` and `evidence/fsc/validation/gradient_check.json`. Only final checkpoints are evaluated.

Before comparing interfaces, each of three $K = 4$ pilot seeds must be tested on eight new 100,000-slot saturated trajectories and eight empty-start trajectories. A seed passes if its mean saturated rate is at least 0.80, its two-sided 95% t-interval lower endpoint exceeds 0.78, its mean final queue is at most 100, and every final queue is at most 1,000. At least two of three must pass. REINFORCE passes $0/3$; its evaluated rates are 0.715385, 0.6970275 and 0.62117375. All 4,500 actual Adam updates and all three failures remain. They identify failure of this budgeted optimizer, not insufficient $K = 4$ capacity or a failure of recurrent RL in general.

The protocol's second route is reward-ranked mutation search over deterministic E, W, T tables. It uses 128 candidates for 160 generations, retaining 16 elites, producing 96 mutated children and 16 independent random immigrants. Every gene starts uniformly and independently; neither action coverage nor a tracker transition is inserted. A mutation replaces $1 + \mathrm{Binomial}(11 + K, 0.08)$ distinct genes by uniform categorical values, which may equal their previous values. Each generation scores every candidate on four 2,048-slot episodes with common environmental tapes. New tapes are drawn at every generation. All final candidates are evaluated on 32 independent 8,192-slot validation episodes, and the largest sampled return selects the frozen controller. Exact rates, boundary kernels, confirmation queues and certificates are unavailable to selection. All three backup pilots pass, after which ten fresh seeds per $K = 3, 4, 8$ are run without retuning.

What is and is not learned. Both the memory transition and action maps are learned from scalar interaction reward, satisfying the joint-learning objective in a finite-controller policy-search sense. The successful route uses deterministic decoder rows, a subclass that restricts the used physical actions when $K < 4$. It does not demonstrate optimization over every randomized row allowed by the theorem. $K = 8$ is a wide-interface capability control, not eight physical actions or evidence of an improvement over $K = 4$. All memory sizes remain $M = 4$. The earlier model-aware FSC search and fixed-recurrence teacher training are distinct experiments, not relabeled instances of this reward-only study. Model-free FSC learning itself is established (Meuleau et al., 1999b).

H.2 Independent continuing evaluation and uncertainty

Every final confirmation controller is evaluated on eight environmental seeds from each of $Q_0 = 0$ and 1000, for 10^6 slots without resets or clipping. Prefixes 10^4 and 10^5 are measured on those same trajectories. Hidden initialization is the environment seed modulo four and internal initialization is zero. The independent scalar evaluator uses MT19937; training uses PCG64. Shared environment seeds are retained across methods to support paired comparisons. They do not create additional independent training replicates.

The primary empirical statistic is mean Q over the last 500,000 slots from $Q_0 = 0$. Uncertainty uses 5,000 crossed bootstrap resamples of the ten training seeds and eight environmental seeds, independently along the two axes and with shared draws across methods. Table 2 gives the percentile 95% intervals. This is a 10×8 crossed design, not 80 independent training runs. Before display rounding, the $K = 3$ late-queue estimate and interval are 46825.278 [46275.802, 47317.689]; for $K = 4$ they are 4.418317 [4.368664, 4.470059]. The hand-coded tracker is separately evaluated as an environment reference and never counted as learned. Final queue, tail fractions, quantiles, last-half regression slope, alternate initialization, and all nested-horizon statistics remain in the machine-readable records. Finite trajectories do not establish positive recurrence.

All $K = 4$ and $K = 8$ confirmations recover the specified tracker up to state and label renaming; their exact saturated rate is $107140/125673$. None of the initial populations contains this tracker.

A first equivalent tracker appears between generations 5 and 60 across the successful pilots and confirmations. The ten $K = 3$ policies have numerically solved saturated rate 0.717237. Their low rates and growing finite queues illustrate the interface obstruction; the class-wide $K = 3$ impossibility follows from the joint bound, independently of these learned policies. Their mean final queue is 62278.175, compared with 4.725 and 4.525 for $K = 4$ and 8. Mean last-half slopes are 0.0621864, -2.77×10^{-7} and -2.89×10^{-7}, respectively. Departure rates of stable empty-start queues approach the arrival rate; the independently computed saturated rate, not that realized departure rate, identifies their service margin.

Table 5: Every confirmation training seed. Entries average the primary late-queue statistic over eight environmental seeds from $Q_0 = 0$. Certificate counts are $0/10$, $10/10$, $10/10$ in the three columns.

Training seed	$K = 3$	$K = 4$	$K = 8$
9100	47076.250	4.414767	4.414767
9101	46629.790	4.418349	4.418349
9102	46629.540	4.414767	4.423018
9103	47076.875	4.418349	4.420454
9104	47146.406	4.423018	4.414767
9105	46771.789	4.418349	4.423018
9106	45857.658	4.423018	4.420454
9107	47146.281	4.414767	4.420454
9108	47146.406	4.414767	4.420454
9109	46771.789	4.423018	4.423018

H.3 Post-training certificates for the learned controllers

For an arbitrary exported E, W, T of the stated architecture, let $i = (s, c)$, $A(c, a) = \sum_z E(z \mid c)W(a \mid z)$ and $r_i = \sum_a A(c, a)R(s, a)$. With $j = (s', c')$, the positive-level blocks are

$$B_-(i, j) = (1 - \lambda)\mathsf{H}(s, s')r_i T(c' \mid c, 1),$$
$$B_+(i, j) = \lambda\mathsf{H}(s, s')(1 - r_i)T(c' \mid c, 0),$$
$$B_0(i, j) = \mathsf{H}(s, s')\{(1 - \lambda)(1 - r_i)T(c' \mid c, 0) + \lambda r_i T(c' \mid c, 1)\}.$$

Write $P = B_- + B_0 + B_+$. At level zero, the actual departure is zero regardless of potential service, so the phase kernel is instead

$$P^{(0)}(i, j) = \mathsf{H}(s, s')T(c' \mid c, 0).$$

Its level-zero and level-one blocks are $(1 - \lambda)P^{(0)}$ and $\lambda P^{(0)}$. These boundary blocks are independently checked.

Proposition 25 (A checked sufficient certificate for the exported FSC)**.** *In the stated finite model, assume $0 < b \le r_i \le g < 1$, $0 < \lambda < 1$, that $P^{(0)}$ is irreducible, and that a finite vector h and $\delta > 0$ satisfy*

$$\lambda - r_i + \sum_j P(i, j)h_j - h_i \le -\delta \qquad \textit{for every } i.$$

Then the joint queue/phase chain is irreducible and positive recurrent. It also has a stationary exponential queue moment.

Proof. Let $V(q, i) = q + h_i - \min_j h_j$. Its conditional drift is at most $-\delta$ at every positive queue level. The level-zero set is finite. From any level, repeated departures without arrivals reach that set with positive probability, since $r_i \ge b > 0$. Its phases communicate by following $P^{(0)}$ while arrivals remain zero. To reach any desired positive level and phase from the boundary, follow a sufficiently long $P^{(0)}$ walk, impose no departures, and place the required number of arrivals along the walk. Every imposed event has positive probability because $r_i \le g < 1$ and $0 < \lambda < 1$. Thus the full chain is irreducible, and finite-set Foster drift gives positive recurrence (Meyn & Tweedie, 1992; 2009).

Put $B = 1 + \max h - \min h$, so $|\Delta V| \le B$. Since $\delta \le B$, choose $\theta = \delta/(3B^2)$, giving $\theta B \le 1$. Taylor's remainder and $e < 3$ imply, outside the boundary,

$$\mathbb{E}[e^{\theta \Delta V} \mid q, i] \le 1 - \theta\delta + \tfrac{3}{2}\theta^2 B^2 = 1 - \tfrac{1}{2}\theta\delta < 1.$$

The finite boundary has bounded exponential drift. The exponential Foster condition yields a finite stationary exponential moment of V, and hence of the queue. □

The verifier interprets each exported floating entry as an exact binary rational and normalizes its row exactly. A numerical linear program only proposes h and a drift margin; the proposal is rationalized on a 10^{-9} grid. All sixteen drift inequalities are then checked with exact fractions, along with exact row sums and the boundary support graph. Deterministic confirmation tables contain only zero and one, avoiding a floating-softmax execution discrepancy. Their worst verified drift is at most -0.07252997858506945. A second implementation explicitly enumerates label, action, service, arrival, hidden advance and memory events and reproduces every saved rational drift, with a separate boundary reachability check. The proof uses standard Foster criteria with exactly verified model-specific drift and boundary conditions.

All twenty adequate-interface confirmations, three successful pilots and the hand tracker pass. The ten undersized confirmations and three failed REINFORCE pilots do not. Failure of this sufficient certificate alone means unverified; for $K = 3$, the existing universal bound separately excludes stability because $443/576 < 39/50$. The gap from a searched policy to the ceiling is not identified as pure optimization error: finite memory, restricted decoder rows and partial observation also matter.

H.4 Interaction budgets and complete run accounting

There are 36 learning runs: three REINFORCE pilots, three population pilots and thirty population confirmations. REINFORCE performs 4,500 actual Adam updates and 73,728,000 environment transitions. Population search performs 5,280 generation updates and no gradient update. Each population run uses 167,772,160 candidate transitions for training and 33,554,432 for final model selection, totaling 201,326,592. Candidate comparisons share environmental tapes: a run generates 1,310,720 training tape slots and 262,144 selection tape slots. These counts must not be presented as independently drawn exogenous samples. Across all learning and selection runs the candidate-transition total is 6,717,505,536. The two routes' budgets differ by a factor of 8.192, precluding a sample-efficiency or equal-budget optimizer comparison.

All initial and final controllers, failed pilots and per-run learning histories are retained. Independent evaluation reconstructs the queue from arrivals and actual departures and checks its empty-boundary behavior. Repeated-seed reproduction is a verification check, not an additional training replication. The anonymous evidence archive includes the original hash-matched trainer, selection and evaluator source, protocol, RNG configuration, initial and final arrays, compact search histories, all 37 controller exports and certificate records. Its default audit checks source provenance, seeded initialization, recorded selection and exact drift certificates. Separate modes rebuild the crossed-bootstrap statistics, replay frozen evaluations, or explicitly rerun training into a new directory. The stored-history audit alone does not re-simulate training rewards.

I Finite-horizon checks of joint recurrent neural learning

This experiment asks whether a learned recurrence and action interface can jointly meet a finite-run queue-competence gate in the theorem's environment. It does not certify GRU stability or propose a generally superior RL algorithm. The primary setting is $m = 4$, $p = 15/16$, $b = 1/4$, $g = 35/36$, $\lambda = 0.78$. Results below use the frozen confirmation protocols; exploratory runs informed these training setups and are retained separately.

I.1 Observations, optimization and deployment

Actor and critic have separate 32-unit GRUs. Their five-dimensional input is the previous action's one-hot encoding and its actual departure, initially zero. Neither receives queue length, hidden phase, potential service, teacher actions or counterfactual rewards. A linear actor head selects a label for $K = 3, 4$, or an action for the unrestricted (Full) comparator. The label-only decoder is a learned $K \times 4$ table of action probabilities. Training uses the exact marginal action likelihood, summing over labels; deployment samples a label and then an action using fresh randomness. Only the selected label and the fixed table are available to the decoder. Recurrence uses action and departure, not the sampled label, so marginalization preserves the contract.

PPO (Schulman et al., 2017) uses 64 parallel trajectories of length 128 per rollout update, four epochs and two sequence minibatches per epoch. Discount and GAE factors are both 0.95; likelihood ratios are clipped to $[0.8, 1.2]$, the entropy coefficient is 0.01, and the gradient norm is clipped at 0.5. Adam uses learning rate 0.0003 and numerical epsilon 10^{-5}. The implementation defines the value loss as half the squared error and weights it by 0.5. The auxiliary setup adds unit-weight binary cross-entropy for a linear head's prediction of the *chosen action's* observed departure. This loss also updates the actor GRU. The decoder-rate intervention changes only the decoder's Adam rate to 0.003. Ordinary PPO sets the auxiliary loss weight to zero. The reward is actual departure D_t. Advantages are normalized jointly over the 128×64 rollout entries. Each rollout bootstraps from its final critic value, including immediately before an administrative reset; the implementation does not apply a terminal-value mask at those resets.

Every 2,048 training steps, the environment and recurrent state reset with $Q_0 = 2049$. Consequently, actual departure equals potential service during training, but only actual departure is returned to the learner. There are no within-episode caps. Empty-queue deployment introduces histories absent from this high-backlog training distribution. At 600 rollout updates, each run uses 4,915,200 interactions and 4,800 Adam steps; 1,200 updates doubles both. The ordinary 600-update checkpoint is a prefix of the 1,200-update run, not another independently trained model. Diagnostic checkpoints do not select the final policy used in the confirmation comparison.

I.2 Complete primary comparison and empirical gate

Ten paired training seeds (34100–34109) are used per setup, with matched initializations for the decoder-rate intervention. Each frozen policy is evaluated for 100,000 slots under eight environment seeds from each of $Q_0 \in \{0, 1000, H_{\text{eval}} + 1\}$, where $H_{\text{eval}} = 100{,}000$. The $H_{\text{eval}} + 1$ starts supply the saturated-rate estimate; the training resets above use a different horizon. The empirical gate requires saturated mean departure at least 0.80, a two-sided 95% t-interval lower endpoint above 0.78, mean empty-start final queue at most 100 and every such final queue at most 1,000. This finite-horizon gate is not a positive-recurrence test. Table 6 retains every primary condition, including the failed adequate-interface setup and the stronger ordinary baseline.

Table 6: Primary stochastic deployment at 100,000 slots. Each row has ten training seeds; queue means use the last half of empty-start trajectories. All entries are empirical, including the "Pass" column.

Training setup (updates)	Finite-run pass	Est. saturated rate	Late mean Q
$K = 4$, auxiliary, default decoder rate (600)	0/10	.659063	9021.585
$K = 4$, auxiliary, faster decoder (600)	10/10	.837010	5.671
$K = 3$, auxiliary, faster decoder (600)	0/10	.704754	5595.027
Full, auxiliary (600)	10/10	.835927	5.764
Full, ordinary (600 prefix)	5/10	.786260	946.626
Full, ordinary (1200)	10/10	.851382	4.417

All ten successful $K = 4$ policies also pass their conditional million-slot stochastic checks, with saturated rate .837433 and late mean queue 5.7635. All 35 eligible long checks pass (ten $K = 4$, ten Full auxiliary, ten ordinary 1,200 and five ordinary 600). These follow-ups are conditional on the primary gate, not additional training seeds or an unconditional long-horizon panel. Conditional on the fixed evaluation bank and gate rule, ten successes give an exact binomial 95% interval $[.6915, 1]$ across training seeds. Increasing the decoder learning rate within the auxiliary $K = 4$ setup gives a paired rate gain of .177947, with training-seed t-interval $[.140557, .215338]$. The insufficient-interface theorem is independent of these outcomes. There is no ordinary-PPO $K = 4$ faster-decoder control, so the necessity of auxiliary prediction for this repair is not established.

I.3 Deployment and stronger-budget controls

Argmax deployment must be distinguished from sampling the trained actor. For Full, the prediction readout selects the action of largest predicted departure. Table 7 evaluates all ten training seeds at one million slots, including unsuccessful ordinary-600 policies. Full readout contrasts pair training seeds, not individual environmental tapes.

Table 7: Full readout controls at one million slots; ten training seeds per row. Large queue means can be dominated by failing seeds.

Training setup and deployment	Est. saturated rate	Late mean Q
Auxiliary 600, actor argmax	.846935	4.8781
Auxiliary 600, prediction argmax	.861405	4.0758
Ordinary 600, actor argmax	.811384	3839.4460
Ordinary 1200, actor argmax	.861098	3.7508

At equal 600-update budgets, the prediction readout improves rate in all ten paired seeds, with mean gain .050021 and 95% interval $[.021877, .078166]$. Ordinary 600 actor argmax passes $7/10$, compared with $5/10$ for sampling. Against ordinary 1,200 actor argmax, auxiliary prediction's rate difference is only .000307 with interval $[-.008647, .009261]$; its queue difference is $+.3250$ with interval $[-.4644, 1.1144]$. These establish neither superiority nor equivalence. Ordinary 1,200 has the lower observed mean queue. The prediction-result files did not define a gate decision; none is imputed.

I.4 Nearby environments and limits of transfer

A frozen replication uses three fresh paired seeds (36101–36103) per environment, auxiliary 600 versus ordinary 1,200 with ordinary-600 prefixes. All use Full heads and load .78. The noncyclic uniform environment has $\mathsf{H}_{ii} = .94$ and $\mathsf{H}_{ij} = .02$ for $i \neq j$, with the original service matrix. The heterogeneous variant keeps this kernel but uses matched rates $(.94, .96, .98, .99)$ and unmatched rate .25. The asymmetric variant retains the original service matrix, with hold probabilities $(.91, .95, .97, .93)$ and otherwise advances cyclically. Appendix G independently certifies an exact joint label-count separation at load .78 for the first two physical models. The asymmetric-persistence model is not a certified joint-separation witness at this load. The theoretical certificates apply to the stated interface and explicit controllers, not to the trained Full-head policies reported here. All 36 specified million-slot readouts are evaluated regardless of their short-horizon gate; eight environmental seeds and all three queue initializations are used. The three tasks are not nine exchangeable replications of one task.

Table 8: Nearby-environment replication: million-slot late mean queue from deterministic actor or prediction argmax readouts of Full-head policies. Each environment has three paired training seeds; all failures are retained. The heterogeneous ordinary-1200 mean includes one failed seed (queue 21564.081) and two near-empty seeds (8.724, 9.841).

Environment	Aux actor	Aux prediction	Ordinary 600	Ordinary 1200
Noncyclic uniform	13.038	7.880	67872.335	7.680
Heterogeneous service	18.581	8.808	86800.637	7194.216
Asymmetric persistence	7.601	10.817	28955.037	11.054

At the fixed 600-update budget, both auxiliary readouts improve rate and late queue over ordinary 600 in every paired seed. Ordinary 1,200 catches up in the uniform environment. As Table 8 shows, one failed seed drives the heterogeneous ordinary-1,200 mean. In the asymmetric environment, prediction deployment worsens its own actor's rate, queue, tail quantile and drainage in all three seeds. Auxiliary stochastic deployment passes $0/3$ short gates in each noncyclic environment and $3/3$ in the asymmetric case; deterministic results cannot be relabeled stochastic-policy success. Earlier zero-shot changes to faster hidden dynamics or noisier service also failed. Thus these paired runs show a descriptive gain over ordinary 600 at the fixed interaction budget in the three specified tasks; they do not establish general RL performance or superiority at the larger budget. Real-valued GRU parameterization does not establish that infinite memory is necessary, and none of these neural policies has an exact stability certificate.

I.5 PER-SEED OUTCOMES AND UNCERTAINTY

Tables 9–12 expose the seed-level queue outcomes and descriptive uncertainty underlying the aggregate comparisons above. These are re-expressions of frozen summaries, not new evaluations. The accompanying CSV also retains each seed's saturated rate and recorded gate status. Intervals across ten training seeds do not address transfer to other environments; the three-seed tasks are shown individually without a pooled significance claim.

Table 9: Primary stochastic deployment: late mean queue for every training seed at 100,000 slots. A star marks a recorded empirical gate pass. K4 default/fast and K3 fast use the auxiliary head; Full aux uses 600 updates, and PPO 600 is a prefix of PPO 1200. These are the six conditions of Table 6; environment repetitions are averaged within each cell.

Seed	K4 default	K4 fast	K3 fast	Full aux	PPO 600	PPO 1200
34100	11509.218	5.804*	5644.280	6.033*	18.596	4.143*
34101	10983.486	5.779*	5497.959	6.249*	9.911*	4.162*
34102	2218.365	5.719*	5582.838	6.050*	14.444*	4.781*
34103	11521.869	5.960*	5315.619	5.292*	4703.398	3.764*
34104	3808.841	5.224*	5552.396	5.532*	8.414*	4.200*
34105	11053.372	6.028*	5910.362	6.078*	16.173	3.749*
34106	11245.101	5.145*	5554.928	5.198*	12.979*	4.151*
34107	11076.886	5.610*	5729.440	5.320*	10.214*	4.192*
34108	4340.445	5.859*	5741.570	6.075*	2340.706	6.171*
34109	12458.269	5.583*	5420.879	5.819*	2331.423	4.852*

Table 10: Full-head deterministic readouts: late mean queue for every training seed at one million slots. Stars mark recorded gate passes; prediction readouts have no recorded gate classification. All ten seeds are shown, including failed ordinary-600 policies.

Seed	Aux actor	Aux prediction	PPO 600	PPO 1200
34100	5.103*	3.988	9.566*	3.488*
34101	5.284*	5.808	6.104*	3.593*
34102	5.222*	3.814	8.532*	3.900*
34103	4.424*	3.475	30182.742	3.172*
34104	4.577*	4.096	5.624*	3.496*
34105	5.254*	4.926	7.717*	3.270*
34106	4.411*	3.338	7.249*	3.665*
34107	4.239*	3.722	5.956*	3.674*
34108	5.289*	3.417	4592.549	5.289*
34109	4.980*	4.175	3568.421	3.961*

Table 11: Nearby-task deterministic readouts: all three training seeds per task, one million slots. Entries are late mean queues; stars mark the recorded finite-run gate. The unaveraged seed values show how failures contribute to Table 8; tasks are not pooled.

Environment	Seed	Aux actor	Aux prediction	PPO 600	PPO 1200
Noncyclic	36101	14.217*	7.220*	41412.109	7.429*
Noncyclic	36102	13.877*	9.082*	96153.890	7.851*
Noncyclic	36103	11.019*	7.337*	66051.007	7.761*
Heterogeneous	36101	13.415*	8.320*	81540.431	8.724*
Heterogeneous	36102	16.437*	8.967*	84800.347	9.841*
Heterogeneous	36103	25.892	9.138*	94061.132	21564.081
Asymmetric	36101	6.290*	9.287*	35539.880	12.951*
Asymmetric	36102	9.150*	12.228*	38114.892	13.975*
Asymmetric	36103	7.363*	10.936*	13210.339	6.236*

Table 12: Means and descriptive 95% Student-t intervals across ten training seeds, conditional on the fixed evaluation procedure. Each seed contributes its environment-averaged statistic. Intervals are not clipped at zero and do not certify recurrence. Primary rows use stochastic deployment; 1M rows use the deterministic Full-head readouts above. PPO 600/1200 share training lineage.

Condition	Saturated rate [95% interval]	Late mean queue [95% interval]
K4 default (100k)	$0.659063\ [0.621907, 0.696219]$	$9021.585\ [6232.961, 11810.209]$
K4 fast (100k)	$0.837010\ [0.835158, 0.838862]$	$5.671\ [5.462, 5.880]$
K3 fast (100k)	$0.704754\ [0.703015, 0.706493]$	$5595.027\ [5472.788, 5717.266]$
Full aux (100k)	$0.835927\ [0.833956, 0.837898]$	$5.764\ [5.484, 6.045]$
PPO 600 (100k)	$0.786260\ [0.760974, 0.811547]$	$946.626\ [-223.411, 2116.663]$
PPO 1200 (100k)	$0.851382\ [0.845696, 0.857068]$	$4.417\ [3.906, 4.927]$
Aux actor (1M)	$0.846935\ [0.844234, 0.849636]$	$4.878\ [4.579, 5.178]$
Aux prediction (1M)	$0.861405\ [0.855991, 0.866819]$	$4.076\ [3.528, 4.624]$
PPO 600 (1M)	$0.811384\ [0.785975, 0.836793]$	$3839.446\ [-2894.077, 10572.969]$
PPO 1200 (1M)	$0.861098\ [0.855328, 0.866868]$	$3.751\ [3.326, 4.175]$