SOLA: SECOND-ORDER LINK ADAPTATION FROM ACK/NACK FEEDBACK

Anonymous authors
Paper under double-blind review

ABSTRACT

Link adaptation is a central control loop in cellular networks: a transmitter must choose a spectral-efficiency target from limited radio-state information and receives only binary ACK/NACK feedback after transmission. We study a contextual threshold-feedback model in which the latent decodable capacity is a linear function of the observed context plus calibrated log-concave fading noise. Under this stationary, model-specified setting, we propose *Second-Order Link Adaptation* (SOLA), an Online-Newton-style method that updates the channel parameter through the ACK/NACK negative log-likelihood and then selects the expected-throughput-maximizing rate. For a fixed unknown channel parameter, known fading CDF, bounded continuous rate set, and adversarially chosen contexts, we prove $\mathcal{O}(d \log T)$ expected effective-throughput regret, including boundary rate optima. Experiments compare SOLA with OLLA, GLM-UCB-style methods, and SALAD, test CDF misspecification, and evaluate discounted D-SOLA under abrupt changes and across 77 recorded 5G context trajectories with synthetic stochastic feedback. These controlled studies support the empirical value of second-order ACK/NACK learning; the theorem does not cover discounting, finite MCS, misspecified CDFs, or safety margins.

1 INTRODUCTION

Link adaptation — the real-time selection of a modulation and coding scheme (MCS) or spectral-efficiency target — is a central control loop in modern cellular networks. Every transport block is shaped by this choice: selecting too conservatively wastes spectrum, while selecting too aggressively can cause decoding failure and trigger Hybrid-ARQ retransmission. This tradeoff is especially consequential in mobility and low-latency settings, where the scheduler often observes only a short sequence of ACK/NACK outcomes before the effective channel state changes. The goal is therefore not simply to maximize an abstract reward, but to learn a rate-selection rule from censored and asymmetric physical-layer feedback.

The deployed workhorse is Outer-Loop Link Adaptation (OLLA) (Sampath et al., 1997; Pedersen et al., 2007), a lightweight controller that adjusts an offset to steer the block error rate (BLER). OLLA is attractive because it is simple, robust, and cheap enough for real-time use. At the same time, its update is largely reactive: it does not directly exploit contextual side information such as CQI, RSRP/RSRQ, Doppler or mobility indicators, load summaries, and recent HARQ outcomes. In fast-changing regimes, a context-oblivious controller may spend many slots re-learning an offset after the channel or traffic regime has already moved on.

A natural alternative is to cast link adaptation as contextual sequential decision making. LinUCB, Thompson sampling, and generalized linear bandit methods (Li et al., 2010; Agrawal & Goyal, 2013; Filippi et al., 2010; Li et al., 2017) provide principled tools for learning from context and feedback. However, the physical-layer problem differs from the standard reward-feedback abstraction in several ways that matter for both modeling and analysis.

Censored threshold feedback. The scheduler does not observe the instantaneous decodable capacity or the margin by which a rate was safe. It observes only whether the selected rate was below a latent threshold. This ACK/NACK signal is informative, but it is one-sided and action-dependent: the same ACK could mean that the rate was nearly optimal or far below capacity, while

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.

the same NACK could result from a small overshoot or a large mismatch. A useful learning rule should therefore use the full likelihood structure of the threshold observation rather than treating ACK/NACK as an ordinary scalar reward.

Asymmetric outage cost. Under-utilization loses only the difference between the selected rate and the feasible rate. Over-estimation can erase the whole transport block and create retransmission overhead. This motivates expected effective throughput, $\rho\,\mathbb{I}\{\mathrm{ACK}\}$, as the basic objective. The present paper models this asymmetry through the throughput objective and the ACK/NACK likelihood, but does not impose an explicit reliability constraint or prove a BLER target guarantee.

Implementation and non-stationarity. Practical systems use finite MCS tables, safety margins, and measurements that may drift because of mobility, interference, scheduler decisions, and beam-management events. These features motivate discounted and conservative implementations. They also make it important to distinguish clean statistical guarantees from empirical stress tests. Our theorem is for a continuous-rate, stationary, known-CDF abstraction; finite-grid effects, parameter drift, and the safe back-off used in trace-driven evaluation are left outside the formal guarantee.

This paper. We formulate contextual link adaptation as a threshold-feedback problem. At each slot, the scheduler observes a context x_t, selects a rate ρ_t, and receives an ACK if the selected rate does not exceed a latent capacity $c_t = x_t^\top \theta^* + N_t$. The residual noise distribution is assumed known and calibrated. Gaussian shadowing on a log scale, Gumbel log-SNR models, and logistic waterfall approximations are motivating examples of log-concave threshold noise. Under the calibrated compact-curvature condition stated below, the ACK/NACK negative log-likelihood is exp-concave in the channel parameter. SOLA exploits this structure with an Online Newton Step update and then greedily selects the rate that maximizes expected throughput under the current estimate.

Our main theoretical contribution is for the *undiscounted* SOLA update in a stationary model: θ^* is fixed, the fading CDF F is known, the rate is chosen from a bounded continuous set, and the comparator is the oracle that knows θ^* and maximizes expected throughput in each context. In this setting, SOLA achieves $\mathcal{O}(d \log T)$ expected throughput regret for arbitrary, possibly adversarial, context sequences. The proof reduces throughput suboptimality to regret on a sequence of exp-concave likelihood losses. The reduction is useful because it separates the statistical difficulty of learning the threshold model from the one-dimensional rate-selection step.

We also study a practical discounted variant, D-SOLA, in which old curvature information is exponentially downweighted. This variant is motivated by non-stationary wireless traces, but its use should be read as an empirical tracking heuristic rather than as part of the logarithmic-regret theorem. Likewise, safety back-off and finite-grid implementation details are included in the experiments, but are not covered by the present regret guarantee.

Summary of Contributions. Our contributions are threefold.

1. **Model and algorithm.** We introduce SOLA, a second-order ACK/NACK learning algorithm for contextual rate adaptation under calibrated log-concave fading models.
2. **Stationary regret guarantee.** We prove $\mathcal{O}(d \log T)$ expected effective-throughput regret for undiscounted SOLA in the fixed-parameter, known-CDF, continuous-rate abstraction. We also include a related regression lower-bound discussion to contextualize the logarithmic dependence.
3. **Empirical evaluation.** We evaluate SOLA/D-SOLA on stationary synthetic channels, abrupt-change stress tests, and an offline 5G trace-driven proxy evaluation. The experiments are intended to assess tracking and goodput behavior without claiming a dynamic-regret or deployment guarantee.

2 Related Works

Online convex optimization and generalized linear bandits. The Online Newton Step (ONS) algorithm of Hazan et al. (2007) achieves logarithmic regret for exp-concave losses; see Hazan (2016) for a textbook treatment. In the bandit setting, generalized linear bandits (Filippi et al., 2010) and related UCB or Thompson-sampling methods (Abbasi-Yadkori et al., 2011; Li et al., 2017; Agrawal

& Goyal, 2013) typically target reward regret through confidence-set exploration. Closely related binary-feedback approaches include OL2M (Zhang et al., 2016), GLOC (Jun et al., 2017), ECOLog (Faury et al., 2022), and self-concordant analyses for logistic bandits (Faury et al., 2020). These works are designed for broad statistical link functions and often use optimism to balance exploration and exploitation. SOLA focuses on a narrower but physically motivated setting: the scheduler is given a calibrated threshold-noise CDF, and the ACK/NACK likelihood induced by that CDF is used directly as an exp-concave surrogate.

Contextual dynamic pricing. Binary purchase feedback has the same threshold structure, with price and revenue corresponding to rate and goodput. The known-noise RMLP method of Javanmard & Nazerzadeh (2019) obtains logarithmic regret using episodic sparse likelihood estimation under i.i.d. contexts and a positive-definite second-moment condition. SOLA instead uses per-slot ONS updates for arbitrary bounded non-anticipating contexts, with no global covariance condition. Its dimension dependence is not a sparsity-adaptive improvement over RMLP. Our bounded rate set also requires explicit treatment of upper-boundary optima. Appendix E gives further context.

Safe, risk-sensitive, and asymmetric exploration. Conservative bandits (Wu et al., 2016; Kazerouni et al., 2017) enforce baseline-relative reward constraints, safe Gaussian-process methods (Sui et al., 2015) restrict actions to certified safe sets, and safe linear or constrained bandits (Amani et al., 2019; Pacchiano et al., 2021; Saxena et al., 2020) maintain confidence sets for reward and constraint models. Risk-averse bandits also study alternatives to mean reward maximization; for example, Sani et al. (2012) formulate a mean–variance risk-return objective. These formulations are complementary to our throughput model. SOLA encodes outage through the effective-throughput reward $\rho\,\mathbb{I}\{\mathrm{ACK}\}$ and the ACK/NACK likelihood, but it does not impose an explicit reliability constraint, certify safe actions, or prove a BLER guarantee. The safe back-off used in trace-driven experiments is therefore an empirical implementation choice rather than a theorem-backed safety mechanism.

Discounting and damping have also appeared in online learning for stability or tracking. For example, damped second-order updates can stabilize portfolio-style OCO problems (Mhammedi & Rakhlin, 2022), and discounted online learning has been studied for drifting environments (Zhang et al., 2024). Our discounted D-SOLA implementation is motivated by the same broad idea of forgetting stale information, but our formal theorem is deliberately limited to the undiscounted stationary update.

Industrial link adaptation and wireless bandits. OLLA (Sampath et al., 1997; Pedersen et al., 2007) and enhanced variants (Blanquez-Casado et al., 2016) remain widely used because of their simplicity and robustness. WiFi rate adaptation heuristics such as SampleRate (Bicket, 2005), RRAA (Wong et al., 2006), and SoftRate (Vutukuru et al., 2009) likewise exploit link feedback but generally do not provide contextual regret guarantees for the model studied here. Formal wireless-bandit work includes unimodal bandits for rate adaptation (Combes & Proutiere, 2014), contextual multi-armed bandits for cellular link adaptation (Saxena et al., 2019), reinforcement-learning formulations for OLLA-style control (Pulliyakode & Kalyani, 2017), and deep reinforcement-learning approaches for radio-resource control (Mismar et al., 2019).

SALAD (Wiesmayr et al., 2025) is particularly close in motivation: it infers SINR from ACK/NACK cross-entropy loss and combines MCS selection with target-BLER feedback control. SOLA instead studies contextual expected-goodput maximization under a calibrated threshold model. Section 6 compares the released SALAD implementation in a common simulation environment and reports both goodput and BLER; this comparison does not reproduce SALAD's over-the-air deployment study.

3 Problem Formulation

We now define the stylized problem for which the regret analysis is stated. The model is intended to isolate the statistical difficulty of learning from ACK/NACK threshold feedback; it is not a full physical-layer simulator. In particular, the theorem below abstracts away HARQ combining, queueing, scheduler coupling across users, finite-blocklength coding details, and discrete MCS-table constraints.

Problem setting. A single transmitter–receiver link operates over transmission opportunities $t = 1, 2, \ldots, T$. At each opportunity, the scheduler observes a bounded context vector $x_t \in \mathcal{X} \subset \mathbb{R}^d$,

selects a spectral-efficiency target ρ_t from a bounded continuous rate set $\mathcal{R} = [0, \rho_{\max}] \subset \mathbb{R}_+$, and receives one-bit feedback $a_t \in \{0, 1\}$. We use *rate* to mean a normalized spectral-efficiency target. A real scheduler may subsequently map this target to an admissible MCS index, but the theorem treats the target as continuous and does not quantify finite-grid rounding effects.

Context vector. The vector x_t may contain receiver-reported and scheduler-side features such as CQI, RSRP/RSRQ, mobility indicators, recent HARQ summaries, and load information. We assume a bounded context space with $R_x := \sup_{x \in \mathcal{X}} \|x\|_2 < \infty$, which is standard after fixed feature normalization or clipping. The context sequence itself need not be stochastic in the regret theorem; it may be chosen arbitrarily, as long as it is non-anticipating with respect to the current channel noise.

Latent capacity threshold. The instantaneous decodable capacity threshold is modeled in log-domain units as

$$c_t = x_t^\top \theta^* + N_t, \qquad \theta^* \in \mathcal{H} \subseteq \{\theta \in \mathbb{R}^d : \|\theta\|_2 \le R_\theta\}, \tag{1}$$

where $\mathcal{H}$ is known, bounded, and convex, and θ^* is a fixed but unknown parameter. We write $\mu_t := x_t^\top \theta^*$ and $R_{\max} := R_x R_\theta$. The theoretical guarantee in Section 5 is for this fixed-parameter model. Parameter drift, abrupt changes, or switching latent regimes are evaluated only empirically.

The log-domain formulation is a convenient way to represent rate or SNR thresholds after normalization. It should not be interpreted as asserting that all physical effects are linear in the raw radio measurements. Rather, the linear predictor is a local statistical abstraction of the predictable part of the capacity threshold. Adding nonlinear features or learned embeddings is possible, but the regret bound is stated for the resulting fixed finite-dimensional feature vector.

Calibrated threshold-noise model. The residual fading or decoder-softness term N_t is drawn i.i.d. from an absolutely continuous distribution with CDF F and density $f = F'$. The scheduler is assumed to know the calibrated threshold-noise CDF F. This is a strong model-specification assumption: misspecified scales, heteroskedastic fading, mixtures of propagation regimes, and uncalibrated BLER curves are not covered by the proof.

Three common physical mechanisms motivate this family of models. First, log-normal shadowing corresponds to Gaussian noise after taking logarithmic or dB-scale quantities. Second, Rayleigh fading can lead to a Gumbel-type log-SNR distribution. Third, finite-blocklength or decoder-waterfall behavior is often locally approximated by a logistic sigmoid. The analysis does not require the noise to be one of these three families; it only uses the structural compact-curvature condition below on the chosen calibrated F.

For compact notation, define

$$\Omega := [-R_{\max} - \rho_{\max},\, R_{\max} + \rho_{\max}], \tag{2}$$

which contains every possible value of $\rho - x^\top \theta$ under the boundedness conditions.

Assumption 1 (Known calibrated threshold-noise CDF with uniform compact curvature)**.** The CDF $F \in C^2(\mathbb{R})$ is strictly increasing on the relevant region, with density f positive on Ω. The functions f and f' are bounded on compact intervals. In addition, on the compact interval Ω,

$$\kappa_c := \inf_{\omega \in \Omega} \min\left\{ -\frac{d^2}{d\omega^2} \log F(\omega), -\frac{d^2}{d\omega^2} \log(1 - F(\omega)) \right\} > 0, \tag{3}$$

and

$$\nu_e := \sup_{\omega \in \Omega} \max\left\{ \frac{f(\omega)^2}{F(\omega)^2}, \frac{f(\omega)^2}{(1 - F(\omega))^2} \right\} < \infty. \tag{4}$$

Assumption 2 (Bounded parameter, context, and rates)**.** The parameter set $\mathcal{H}$ is closed, bounded, and convex, $\|\theta\|_2 \le R_\theta$ for all $\theta \in \mathcal{H}$, the context set satisfies $\|x\|_2 \le R_x$ for all $x \in \mathcal{X}$, and all chosen rates lie in $\mathcal{R} = [0, \rho_{\max}]$. Consequently, $|x^\top \theta| \le R_{\max}$ for all $x \in \mathcal{X}$ and $\theta \in \mathcal{H}$.

Assumption 1 states the positive compact curvature used in the proof explicitly. Strict log-concavity of F and $1 - F$ is a useful modeling guide, but the proof requires the displayed uniform lower bound on Ω rather than strict log-concavity alone.

ACK/NACK threshold feedback. A codeword transmitted at rate ρ_t is decoded successfully if and only if it does not exceed the instantaneous threshold:

$$a_t = \mathbb{I}\{\rho_t \le c_t\} = \mathbb{I}\{\rho_t - x_t^\top \theta^* \le N_t\}. \tag{5}$$

The exact margin $c_t - \rho_t$ is not observed; the scheduler only sees (x_t, ρ_t, a_t). This censored observation is the sole feedback used by SOLA.

Objective: effective throughput. The delivered goodput in slot t is $r_t = \rho_t a_t$. Conditional on (x_t, ρ), its expectation is

$$\Gamma(\rho, \mu) := \mathbb{E}[r_t \mid x_t, \rho] = \rho\big(1 - F(\rho - \mu)\big). \tag{6}$$

The factor $1 - F(\rho - \mu)$ is the success probability, and the multiplicative rate ρ captures the cost of being conservative. Thus the objective naturally incorporates the outage-induced collapse of delivered throughput, although it is not a constrained reliability objective.

For each predicted mean μ, define the continuous-rate oracle map

$$\rho^*(\mu) \in \arg\max_{\rho \in \mathcal{R}} \Gamma(\rho, \mu). \tag{7}$$

When the maximizer is not unique, any fixed deterministic tie-breaking rule may be used. The compact rate set ensures existence of a maximizer and keeps all constants used in the analysis finite.

Regret formulation. A policy A maps past observations and the current context to a rate. We compare it with the oracle scheduler that knows θ^* and uses $\rho^*(\mu_t)$ in each slot. The cumulative expected-throughput regret is

$$\mathrm{Reg}_T(A) := \sum_{t=1}^{T} \left[\Gamma(\rho^*(\mu_t), \mu_t) - \Gamma(A(x_t, \mathrm{hist}(t-1)), \mu_t)\right]. \tag{8}$$

The context sequence may be adversarial or adaptive, but the parameter θ^* and the noise law remain fixed in the theorem.

4 Algorithm: Second-Order Link Adaptation (SOLA)

SOLA maintains a channel-parameter estimate using the negative log-likelihood induced by ACK/NACK feedback. Given the current estimate, it chooses the rate that maximizes the model-predicted effective throughput. The algorithm stated below is the undiscounted SOLA update analyzed in Section 5; discounted and conservative variants used in the empirical section are described separately after the algorithm.

Likelihood loss. For the selected rate ρ_t and observed ACK/NACK a_t, the negative log-likelihood is

$$\ell_t(\theta) = -a_t \log\big(1 - F(\rho_t - x_t^\top \theta)\big) - (1 - a_t) \log F(\rho_t - x_t^\top \theta). \tag{9}$$

This loss uses only the censored feedback and the calibrated CDF F; the latent capacity c_t is never observed. Once (x_t, ρ_t, a_t) is realized, $\ell_t(\cdot)$ is a convex exp-concave function over $\mathcal{H}$ under Assumption 1 and Assumption 2. The loss is adaptive because ρ_t is selected by the learner, but it is non-anticipating: the rate is chosen before a_t is observed.

Rate selection. At the beginning of slot t, SOLA forms the predicted mean $\widehat{\mu}_t = x_t^\top \theta_t$ and computes

$$\widehat{\rho}_t \in \arg\max_{\rho \in \mathcal{R}} \rho\big(1 - F(\rho - \widehat{\mu}_t)\big). \tag{10}$$

This is a one-dimensional optimization problem. In implementation it can be solved by a bounded scalar optimizer or by a dense rate grid. The theorem requires the continuous-rate maximizer; the experiments use grid-based implementations matching the plotted artifacts.

Second-order update. The matrix A_t is the standard ONS outer-product metric for exp-concave likelihood losses; it is not a Hessian estimate of the throughput objective. The likelihood curvature in Lemma 4 is what allows this metric to control the prediction error in the direction x_t.

Algorithm 1: SOLA: analyzed undiscounted update

Input: Parameter set $\mathcal{H} \subset \mathbb{R}^d$; rate set $\mathcal{R} = [0, \rho_{\max}]$; calibrated CDF F; ONS step size η; regularizer $\lambda_0 > 0$.

Initialize $\theta_1 \in \mathcal{H}$ and $A_0 = \lambda_0 I_d$;
for $t = 1, 2, \ldots, T$ **do**
  Observe context x_t;
  $\widehat{\mu}_t \leftarrow x_t^\top \theta_t$;
  $\rho_t \leftarrow \arg\max_{\rho \in \mathcal{R}} \rho\big(1 - F(\rho - \widehat{\mu}_t)\big)$;
  Transmit at rate ρ_t and observe $a_t \in \{0, 1\}$;
  Form ℓ_t as in (9) and set $g_t \leftarrow \nabla \ell_t(\theta_t)$;
  $A_t \leftarrow A_{t-1} + g_t g_t^\top$;
  $\theta_{t+1} \leftarrow \Pi_{\mathcal{H}}^{A_t}\big(\theta_t - \eta A_t^{-1} g_t\big)$;

D-SOLA implementation used in non-stationary experiments. The non-stationary and trace-driven experiments use a practical variant, D-SOLA, that exponentially discounts older curvature information. In inverse form, with discount $\delta \in (0, 1]$, the implemented Sherman–Morrison update is

$$u_t = A_{t-1}^{-1} g_t, \qquad A_t^{-1} = \frac{1}{\delta}\left(A_{t-1}^{-1} - \frac{u_t u_t^\top}{\delta + g_t^\top u_t}\right), \tag{11}$$

$$\theta_{t+1} = \Pi_{\mathcal{H}}\big(\theta_t - \eta A_t^{-1} g_t\big). \tag{12}$$

This discounts the accumulated outer-product metric while retaining the exact likelihood gradient. The implementation uses Euclidean projection; the ridge and denominator check are numerical safeguards. Trace-context experiments additionally apply a fixed back-off margin to the greedy grid rate. Restarting the analyzed SOLA update at known stationary-segment boundaries gives a rigorous benchmark for discarding stale information (Appendix G); it does not prove a bound for the discounted update above, its projection, or safety margins.

Remark 1 (Computational overhead). The dense inverse-metric update costs $O(d^2)$ per slot, in addition to rate selection and projection. A single-thread Python benchmark measures median total times of $19.87\,\mu$s at $d = 10$ and $45.46\,\mu$s at $d = 100$ (Appendix G). These are offline implementation timings, not production-scheduler latency guarantees.

5 Theoretical Analysis

This section states the regret guarantee for the algorithm actually analyzed in the paper: undiscounted SOLA as written in Algorithm 1. Discounting, empirical back-off, parameter drift, and finite-MCS rounding are not part of the theorem.

Theorem 2 (Stationary regret of SOLA)**.** *Suppose Assumption 1 and 2 hold. Assume θ^* is fixed, N_t are i.i.d. with known calibrated CDF F, the rate set is the bounded continuous set $\mathcal{R}$, and SOLA is run as in Algorithm 1. Let*

$$D_\theta := \sup_{u,v \in \mathcal{H}} \|u - v\|_2, \qquad G_\ell := R_x \sqrt{\nu_e}, \qquad \alpha := \frac{\kappa_c}{\nu_e},$$

and, when $D_\theta G_\ell > 0$, define

$$\beta := \frac{1}{2} \min\left\{\alpha, \frac{1}{4 D_\theta G_\ell}\right\}. \tag{13}$$

Run ONS with $A_0 = \lambda_0 I_d$, $\lambda_0 > 0$, and an admissible ONS step size $\eta \geq 1/\beta$; the canonical choice in the proof is $\eta = 1/\beta$. Then for any bounded, non-anticipating context sequence $\{x_t\}_{t=1}^T$, there exist constants $C_0, C_1 < \infty$ depending on $R_x, R_\theta, \rho_{\max}, \lambda_0, \eta, \kappa_c, \nu_e$, and the calibrated CDF F on Ω, but not on T, such that

$$\mathbb{E}[\mathrm{Reg}_T(\mathrm{SOLA})] \leq C_0 + C_1 d \log\left(1 + \frac{T}{d}\right). \tag{14}$$

The comparator is the oracle that knows θ^* and chooses $\rho^*(x_t^\top \theta^*)$ in each slot.

The proof has two ingredients. First, regret in expected throughput is locally controlled by the prediction error of the mean threshold. Under the calibrated compact-curvature assumptions, the one-dimensional throughput maximizer is monotone and one-Lipschitz in the predicted mean on the relevant compact interval; the appendix proves this statement using the hazard-rate characterization and explicitly handles the boundary case.

Lemma 3 (Throughput-regret upper bound)**.** *Under Assumption 1 and 2, there exists $M_\Gamma < \infty$ depending only on $\rho_{\max}$ and the calibrated CDF F on Ω such that, for every $\theta \in \mathcal{H}$ and $x \in \mathcal{X}$,*

$$\Gamma(\rho^*(x^\top\theta^*), x^\top\theta^*) - \Gamma(\rho^*(x^\top\theta), x^\top\theta^*) \le M_\Gamma(\theta-\theta^*)^\top xx^\top(\theta-\theta^*). \tag{15}$$

Second, the ACK/NACK likelihood has curvature in exactly the same direction. For a fixed realized tuple (x_t, ρ_t, a_t), the loss $\ell_t(\theta)$ is a convex exp-concave function of θ.

Lemma 4 (Likelihood curvature)**.** *Under Assumption 1 and 2, for every realized (x_t, ρ_t, a_t) and every $\theta \in \mathcal{H}$,*

$$\nabla^2\ell_t(\theta) \succeq \kappa_c x_t x_t^\top, \qquad \nabla^2\ell_t(\theta) \succeq \frac{\kappa_c}{\nu_e}\nabla\ell_t(\theta)\nabla\ell_t(\theta)^\top. \tag{16}$$

The first inequality says that, conditional on the selected rate, the likelihood has curvature along the observed context direction. The second inequality is the exp-concavity-type condition needed by ONS. Both constants are finite because all relevant values of $\rho - x^\top\theta$ lie in the compact interval Ω, and the positive curvature lower bound is stated explicitly in Assumption 1.

Adaptive losses. A minor technical point is that ℓ_t depends on the learner's rate ρ_t. This does not invalidate the OCO step. The action ρ_t is chosen before a_t is observed; once (x_t, ρ_t, a_t) is realized, ONS receives a pathwise convex exp-concave loss $\ell_t(\cdot)$. Standard ONS regret bounds apply to such non-anticipating adaptive loss sequences. Moreover, the conditional population likelihood is minimized at the data-generating parameter θ^*, so the likelihood regret controls the squared prediction error that appears in Lemma 3.

Combining Lemma 3 and 4 with the ONS regret bound gives Theorem 2. The detailed proof, including the dependence of the constants on $\lambda_0, \eta, \kappa_c, \nu_e$ and the compactness parameters, is in Appendix C. The theorem is intended to establish the logarithmic growth rate under the model assumptions, not to certify the discounted or finite-grid implementations used in later experiments.

Related lower-bound benchmark. Appendix D recalls an online-regression lower bound of order $\Omega(d\log(1+T/d))$. We use it only as a benchmark for the statistical scale of second-order estimation. It is not a complete minimax lower bound for the censored throughput-feedback problem with all wireless implementation details. For this reason, the main text avoids claiming information-theoretic optimality for the full link-adaptation problem.

6 Numerical Experiments

We first assess SOLA's learning behavior in stationary channels with a known noise CDF. We then examine its empirical robustness and tracking through CDF misspecification, finite-MCS comparisons, abrupt channel changes, and trace-context simulations.

Theory–experiment mapping. The stationary known-CDF experiment is a sanity check for the model in Theorem 2, rather than a verification of its continuous maximization and ONS parameter conditions. Abrupt changes test D-SOLA's tracking; recorded contexts with synthetic feedback test session variability. CDF misspecification and finite-MCS comparisons probe additional modeling choices. These stress tests provide no dynamic-regret or deployment guarantee.

Baselines, metrics, and settings. We compare with OLLA, GLM-UCB-style methods, and released SALAD (Wiesmayr et al., 2025). Regret uses expected goodput against the relevant grid oracle; realized goodput is $\rho_t a_t$ and outage/BLER is $1 - a_t$. Synthetic SOLA/GLM-UCB settings are fixed across seeds and noise families. SALAD uses constructor defaults except for the rate table and target BLER, without knowledge distillation. No CDF condition is retuned; trace margins are global. Appendix F gives protocols and aggregation rules.

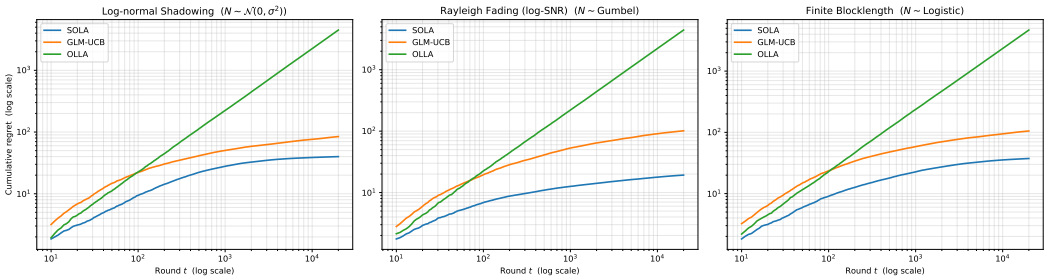

Figure 1: **Stationary known-CDF sanity check.** Cumulative expected effective-throughput regret against the grid oracle under Gaussian, Gumbel, and logistic noise. Runs use $T = 20{,}000$, $d = 10$, normalized bias-augmented contexts, and rates $0.05, 0.10, \ldots, 4.0$. Curves average ten runs, without variability bands. For log-log display, the original plotting routine shows $t \geq 10$ and floors plotted means at 10^{-3}. This finite-grid implementation does not directly verify the continuous-rate theorem.

Stationary Synthetic Channels. We use $T = 20{,}000$, $d = 10$, and Gaussian, Gumbel, or logistic noise calibrated to marginal standard deviation about 0.3. In Figure 1, SOLA's cumulative expected-throughput regret grows more slowly than the OLLA and GLM-UCB-style baselines. This matched-model comparison is separate from the calibration stress test below.

CDF Misspecification. We test 22 conditions with ten paired seeds and a true-CDF grid oracle. With true Gumbel noise and an assumed logistic CDF, mean expected goodput changes from 0.4043 to 0.4003: paired change -0.0040, 95% CI $[-0.0042, -0.0038]$. Across four mean/variance-matched family mismatches, the largest point-estimate loss is about 1.2%. With $\pm 25\%$ same-family scale error, retained goodput is 96.7%–99.8%; larger scale errors degrade performance more. These observations support tolerance to moderate errors, not arbitrary misspecification.

Finite-MCS Comparison with SALAD. An environment adapter gives all methods the same 30-MCS table, logistic BLER curves, ACK convention, and one-slot feedback, without modifying SALAD's estimator or control logic. Across ten paired seeds, SOLA obtains goodput 1.613 ± 0.007 and BLER 0.0783 ± 0.0026, versus SALAD's 1.454 ± 0.025 and 0.1004 ± 0.0004 (mean $\pm$ seed SD). The paired goodput gain is 0.159 with 95% CI $[0.145, 0.172]$. SALAD regulates target BLER 0.10, whereas SOLA maximizes model-based goodput; both metrics are therefore reported. This is a common-environment comparison, not an over-the-air replication.

Abrupt-Change Stress Test. In a Gaussian environment with $T = 30{,}000$ and $d = 10$, the first change flips non-bias parameter signs at $t = 10{,}000$; the second redraws the parameter at $t = 20{,}000$. Figure 2 shows that D-SOLA recovers a lower-error regime than D-GLM-UCB and accumulates regret more slowly between changes. The result measures empirical tracking, without a dynamic-regret or fixed recovery-time claim.

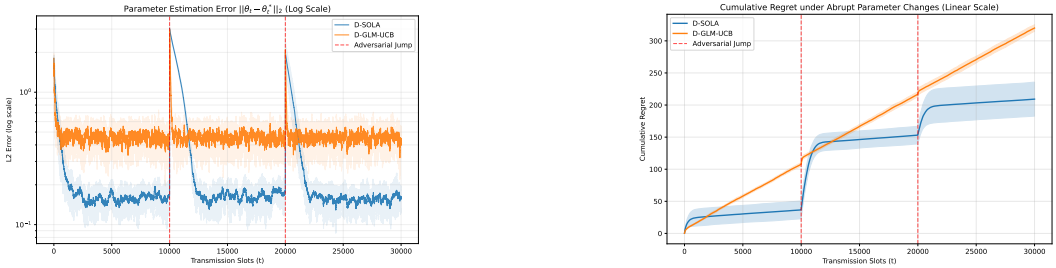

Figure 2: **Abrupt-change stress test.** *Left:* pre-update parameter error $\|\theta_t - \theta_t^*\|_2$. *Right:* cumulative expected regret against the current segment's grid oracle. Changes at $t = 10{,}000$ and $t = 20{,}000$ test empirical tracking outside the stationary theorem. Curves and bands show mean ± 1 SD over ten runs (divisor ten). D-SOLA and D-GLM-UCB use discounts 0.998 and 0.995, respectively.

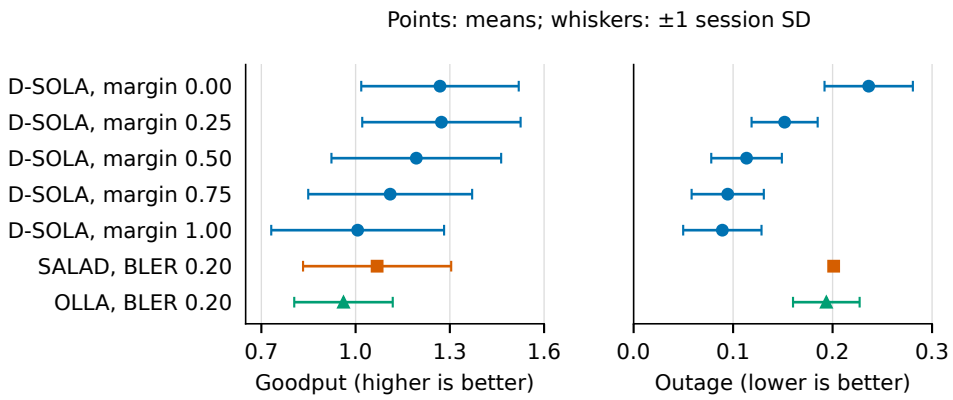

Figure 3: **Controlled 5G trace-context simulation.** Mean goodput (left) and outage (right) for all five global D-SOLA margins and SALAD/OLLA (target BLER 0.20). Recorded contexts drive fixed per-session linear thresholds with matched-logistic stochastic feedback. Ten common-random-number seeds are averaged per session; points and bars show mean ± 1 empirical SD across 77 session means, not confidence intervals. This tests context variability, not within-session drift or deployment performance.

Controlled 5G Trace-Context Simulation. We evaluate 77 retained sessions from the public *5G-Production-Dataset*[1]. CQI/SNR/RSRP standardization and fitting a fixed linear latent threshold to smoothed DL_bitrate use each entire session; ACK/NACK feedback is then sampled with matched-logistic noise. This is an offline construction of a controlled environment, not online train/test evaluation or measured counterfactual decoding. Ten common-random-number seeds are averaged per session, then sessions receive equal weight.

Figure 3 reports all five global D-SOLA margins alongside SALAD and OLLA, both targeting BLER 0.20. Descriptively, margin 0.25 has the highest observed mean goodput, 1.273, with outage 0.152, versus SALAD's 1.069 and 0.201. The paired goodput gain is 0.205 (95% CI $[0.161, 0.245]$), with outage reduction 0.049 (CI $[0.042, 0.056]$). These pointwise CIs resample whole sessions 4,000 times; figure bars show session SD. The highlighted margin is a descriptive choice, not a selection-adjusted optimum.

7 Discussions

SOLA shows that calibrated physical fading models can be useful not only for link-budget calculations, but also for online learning from ACK/NACK threshold feedback. Under a fixed contextual capacity parameter, known calibrated threshold-noise CDF with compact curvature, and bounded continuous rate set, the ACK/NACK likelihood is exp-concave and supports a simple Online-Newton-style update with $O(d \log T)$ expected throughput regret. The empirical results suggest that this curvature-adaptive update can be competitive with OLLA and GLM-UCB-style baselines in stationary simulations, abrupt-change stress tests, and an offline trace-driven proxy evaluation.

Limitations and Extensions. The result is deliberately narrower than a full wireless deployment problem. First, the theorem assumes a fixed θ^* and i.i.d. calibrated noise; the discounted D-SOLA implementation used in stress tests is not covered by the logarithmic-regret proof. Second, the analysis uses a continuous bounded rate abstraction and does not quantify finite-MCS rounding effects. Third, the objective is expected effective throughput rather than a constrained-reliability or tail-risk objective, so the empirical safe back-off in the trace study has no formal BLER guarantee. Finally, the trace-context evaluation uses fixed per-session linear thresholds with synthetic stochastic feedback and should not be treated as deployment validation. See Section G for more discussions.

[1]Public traces: `https://github.com/acmmmsys/2020-5Gdataset`, released under GPL-3.0.

AI USE STATEMENT

Generative AI tools assisted with experimental code and analysis of the results. They were also used for figure preparation and manuscript editing. These tools supported the research and writing workflow under the author's direction. The author takes responsibility for the accuracy and integrity of the final content.

REPRODUCIBILITY STATEMENT

The model assumptions and algorithm are specified in Sections 3–4, with supporting proofs in Appendices A–D. Section 6 and Appendix F document the evaluation protocols, parameter settings, and data preprocessing. An anonymized Python package implements the original synthetic experiments; the public 5G traces are obtained separately.

REFERENCES

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24, 2011.

Shipra Agrawal and Navin Goyal. Thompson sampling for contextual bandits with linear payoffs. In *International conference on machine learning*, pp. 127–135. PMLR, 2013.

Sanae Amani, Mahnoosh Alizadeh, and Christos Thrampoulidis. Linear stochastic bandits under safety constraints. *Advances in Neural Information Processing Systems*, 32, 2019.

John Charles Bicket. *Bit-rate selection in wireless networks*. PhD thesis, Massachusetts Institute of Technology, 2005.

Francisco Blanquez-Casado, Gerardo Gomez, Maria del Carmen Aguayo-Torres, and Jose Tomas Entrambasaguas. eolla: an enhanced outer loop link adaptation for cellular networks. *EURASIP Journal on Wireless Communications and Networking*, 2016(1):20, 2016.

Maxime C Cohen, Ilan Lobel, and Renato Paes Leme. Feature-based dynamic pricing. *Management Science*, 66(11):4921–4943, 2020.

Richard Combes and Alexandre Proutiere. Unimodal bandits: Regret lower bounds and optimal algorithms. In *International Conference on Machine Learning*, pp. 521–529. PMLR, 2014.

Louis Faury, Marc Abeille, Clément Calauzènes, and Olivier Fercoq. Improved optimistic algorithms for logistic bandits. In *International Conference on Machine Learning*, pp. 3052–3060. PMLR, 2020.

Louis Faury, Marc Abeille, Kwang-Sung Jun, and Clément Calauzènes. Jointly efficient and optimal algorithms for logistic bandits. In *International Conference on Artificial Intelligence and Statistics*, pp. 546–580. PMLR, 2022.

Sarah Filippi, Olivier Cappe, Aurélien Garivier, and Csaba Szepesvári. Parametric bandits: The generalized linear case. *Advances in neural information processing systems*, 23, 2010.

Elad Hazan. Introduction to online convex optimization. *Foundations and Trends in Optimization*, 2 (3-4):157–325, 2016.

Elad Hazan, Amit Agarwal, and Satyen Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2):169–192, 2007.

Adel Javanmard and Hamid Nazerzadeh. Dynamic pricing in high-dimensions. *The Journal of Machine Learning Research*, 20(1):315–363, 2019.

Kwang-Sung Jun, Aniruddha Bhargava, Robert Nowak, and Rebecca Willett. Scalable generalized linear bandits: Online computation and hashing. *Advances in Neural Information Processing Systems*, 30, 2017.

Abbas Kazerouni, Mohammad Ghavamzadeh, Yasin Abbasi Yadkori, and Benjamin Van Roy. Conservative contextual linear bandits. *Advances in neural information processing systems*, 30, 2017.

Robert Kleinberg and Tom Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *IEEE Symposium on Foundations of Computer Science (FOCS-03)*, pp. 594–605. IEEE, 2003.

Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pp. 661–670, 2010.

Lihong Li, Yu Lu, and Dengyong Zhou. Provably optimal algorithms for generalized linear contextual bandits. In *International Conference on Machine Learning*, pp. 2071–2080. PMLR, 2017.

Zakaria Mhammedi and Alexander Rakhlin. Damped online newton step for portfolio selection. In *Conference on learning theory*, pp. 5561–5595. PMLR, 2022.

Faris B Mismar, Brian L Evans, and Ahmed Alkhateeb. Deep reinforcement learning for 5g networks: Joint beamforming, power control, and interference coordination. *IEEE Transactions on Communications*, 68(3):1581–1592, 2019.

Aldo Pacchiano, Mohammad Ghavamzadeh, Peter Bartlett, and Heinrich Jiang. Stochastic bandits with linear constraints. In *International conference on artificial intelligence and statistics*, pp. 2827–2835. PMLR, 2021.

Klaus I Pedersen, Guillaume Monghal, Istvan Z Kovacs, Troels E Kolding, Akhilesh Pokhariyal, Frank Frederiksen, and Preben Mogensen. Frequency domain scheduling for ofdma with limited and noisy channel feedback. In *2007 IEEE 66th Vehicular Technology Conference*, pp. 1792–1796. IEEE, 2007.

Saishankar Katri Pulliyakode and Sheetal Kalyani. Reinforcement learning techniques for outer loop link adaptation in 4g/5g systems. *arXiv preprint arXiv:1708.00994*, 2017.

Ashwin Sampath, P Sarath Kumar, and Jack M Holtzman. On setting reverse link target sir in a cdma system. In *1997 IEEE 47th Vehicular Technology Conference. Technology in Motion*, volume 2, pp. 929–933. IEEE, 1997.

Amir Sani, Alessandro Lazaric, and Rémi Munos. Risk-aversion in multi-armed bandits. *Advances in neural information processing systems*, 25, 2012.

Vidit Saxena, Joakim Jaldén, Joseph E Gonzalez, Mats Bengtsson, Hugo Tullberg, and Ion Stoica. Contextual multi-armed bandits for link adaptation in cellular networks. In *Proceedings of the 2019 Workshop on Network Meets AI & ML*, pp. 44–49, 2019.

Vidit Saxena, Joakim Jalden, and Joseph Gonzalez. Thompson sampling for linearly constrained bandits. In *International Conference on Artificial Intelligence and Statistics*, pp. 1999–2009. PMLR, 2020.

Yanan Sui, Alkis Gotovos, Joel Burdick, and Andreas Krause. Safe exploration for optimization with gaussian processes. In *International conference on machine learning*, pp. 997–1005. PMLR, 2015.

Mythili Vutukuru, Hari Balakrishnan, and Kyle Jamieson. Cross-layer wireless bit rate adaptation. In *Proceedings of the ACM SIGCOMM 2009 conference on Data communication*, pp. 3–14, 2009.

Reinhard Wiesmayr, Lorenzo Maggi, Sebastian Cammerer, Jakob Hoydis, Fayçal Aït Aoudia, and Alexander Keller. Salad: Self-adaptive link adaptation. *arXiv preprint arXiv:2510.05784*, 2025.

Starsky HY Wong, Hao Yang, Songwu Lu, and Vaduvur Bharghavan. Robust rate adaptation for 802.11 wireless networks. In *Proceedings of the 12th annual international conference on Mobile computing and networking*, pp. 146–157, 2006.

Yifan Wu, Roshan Shariff, Tor Lattimore, and Csaba Szepesvári. Conservative bandits. In *International Conference on Machine Learning*, pp. 1254–1262. PMLR, 2016.

Lijun Zhang, Tianbao Yang, Rong Jin, Yichi Xiao, and Zhi-Hua Zhou. Online stochastic linear optimization under one-bit feedback. In *International Conference on Machine Learning*, pp. 392–401. PMLR, 2016.

Zhiyu Zhang, David Bombara, and Heng Yang. Discounted adaptive online learning: Towards better regularization. *arXiv preprint arXiv:2402.02720*, 2024.

APPENDIX: PROOFS AND SUPPLEMENTARY DETAILS FOR SECOND-ORDER LINK ADAPTATION (SOLA)

This appendix gives proof details for the stationary SOLA theorem. Throughout this appendix, SOLA means Algorithm 1 with discount $\gamma = 1$ and no empirical safe back-off. The discounted and trace-specific variants used in experiments are not analyzed here.

For reference, write

$$I_\mu := [-R_{\max}, R_{\max}], \qquad \Omega := [-R_{\max} - \rho_{\max},\ R_{\max} + \rho_{\max}],$$

so that $x^\top\theta \in I_\mu$ and $\rho - x^\top\theta \in \Omega$ for all $x \in \mathcal{X}$, $\theta \in \mathcal{H}$, and $\rho \in \mathcal{R}$.

A PROOF OF THE THROUGHPUT REGRET REDUCTION (LEMMA 3)

This section proves Lemma 3 with an explicit coefficient. Let

$$S(\omega) := 1 - F(\omega), \qquad h(\omega) := \frac{f(\omega)}{S(\omega)}.$$

Here h is the hazard rate of the residual noise. Since S is strictly log-concave, $h'(\omega) = -(\log S)''(\omega) \geq 0$. On the compact interval Ω, define

$$H_0 := \sup_{\omega\in\Omega} h(\omega) = \sup_{\omega\in\Omega} \frac{f(\omega)}{1 - F(\omega)}, \tag{17}$$

$$H_1 := \sup_{\omega\in\Omega} h'(\omega) = \sup_{\omega\in\Omega} \frac{f'(\omega)(1 - F(\omega)) + f(\omega)^2}{(1 - F(\omega))^2}. \tag{18}$$

Both quantities are finite: S is bounded away from zero on Ω, and Assumption 1 bounds f and f' on compact intervals. The coefficient used below is

$$M_\Gamma := \frac{1}{2}H_0 + \frac{3}{2}\rho_{\max}H_1 = \frac{1}{2}\sup_{\omega\in\Omega} \frac{f(\omega)}{1 - F(\omega)} + \frac{3\rho_{\max}}{2}\sup_{\omega\in\Omega} \frac{f'(\omega)(1 - F(\omega)) + f(\omega)^2}{(1 - F(\omega))^2}. \tag{19}$$

This expression depends only on the bounded interval determined by $R_x, R_\theta, \rho_{\max}$ and on the calibrated CDF F.

Fix a mean value $\mu \in I_\mu$ and define

$$q_\mu(\rho) := \rho h(\rho - \mu), \qquad \rho \in [0, \rho_{\max}].$$

The derivative of the expected throughput with respect to the selected rate is

$$\frac{\partial}{\partial\rho}\Gamma(\rho, \mu) = S(\rho - \mu) - \rho f(\rho - \mu) = S(\rho - \mu)\big(1 - q_\mu(\rho)\big). \tag{20}$$

Moreover,

$$\frac{\partial}{\partial\rho}q_\mu(\rho) = h(\rho - \mu) + \rho h'(\rho - \mu) \geq 0.$$

Thus $\rho \mapsto q_\mu(\rho)$ is nondecreasing, $q_\mu(0) = 0$, and the maximizer of $\Gamma(\rho, \mu)$ over $[0, \rho_{\max}]$ is characterized as follows: if $q_\mu(\rho_{\max}) \leq 1$, then $\rho^*(\mu) = \rho_{\max}$; otherwise $\rho^*(\mu)$ is the unique point in $(0, \rho_{\max})$ satisfying $q_\mu(\rho^*(\mu)) = 1$.

We first record a Lipschitz property of the oracle rate map. Let $\mu_2 \geq \mu_1$ and $\Delta = \mu_2 - \mu_1$. Since h is nondecreasing,

$$q_{\mu_2}(\rho^*(\mu_1)) = \rho^*(\mu_1)h(\rho^*(\mu_1) - \mu_2) \leq \rho^*(\mu_1)h(\rho^*(\mu_1) - \mu_1) \leq 1,$$

so the optimum under μ_2 cannot be smaller than $\rho^*(\mu_1)$. If $\rho^*(\mu_1) = \rho_{\max}$ this already gives equality. Otherwise $q_{\mu_1}(\rho^*(\mu_1)) = 1$. If $\rho^*(\mu_1) + \Delta \geq \rho_{\max}$, then $\rho^*(\mu_2) \leq \rho_{\max} \leq \rho^*(\mu_1) + \Delta$. If $\rho^*(\mu_1) + \Delta < \rho_{\max}$, then

$$q_{\mu_2}(\rho^*(\mu_1) + \Delta) = (\rho^*(\mu_1) + \Delta)h(\rho^*(\mu_1) - \mu_1) \geq q_{\mu_1}(\rho^*(\mu_1)) = 1,$$

so the maximizer under μ_2 is at most $\rho^*(\mu_1) + \Delta$. Therefore $\rho^*(\mu_2) \le \rho^*(\mu_1) + \Delta$. We have proved

$$0 \le \rho^*(\mu_2) - \rho^*(\mu_1) \le \mu_2 - \mu_1 \qquad \text{whenever } \mu_2 \ge \mu_1. \tag{21}$$

In particular, ρ^* is nondecreasing and one-Lipschitz on I_μ.

Now fix two means $\mu_0, \mu_1 \in I_\mu$ and write

$$r_0 := \rho^*(\mu_0), \qquad r_1 := \rho^*(\mu_1), \qquad \Delta := |\mu_0 - \mu_1|, \qquad \delta := |r_0 - r_1|.$$

By (21), $\delta \le \Delta$. We prove

$$\Gamma(r_0, \mu_0) - \Gamma(r_1, \mu_0) \le M_\Gamma(\mu_0 - \mu_1)^2. \tag{22}$$

If $r_0 = r_1$, the claim is immediate. Suppose first that $r_1 < r_0$. Then $\mu_1 < \mu_0$, and r_1 cannot be the upper boundary, so $q_{\mu_1}(r_1) = 1$. For any $\rho \in [r_1, r_0]$, the characterization of r_0 gives $q_{\mu_0}(\rho) \le 1$. Using (20), $S \le 1$, and the derivative bounds

$$\left|\frac{\partial q_\mu(\rho)}{\partial \rho}\right| \le H_0 + \rho_{\max} H_1, \qquad \left|\frac{\partial q_\mu(\rho)}{\partial \mu}\right| \le \rho_{\max} H_1,$$

we obtain

$$\begin{aligned}
\Gamma(r_0, \mu_0) - \Gamma(r_1, \mu_0) &= \int_{r_1}^{r_0} S(\rho - \mu_0)\big(1 - q_{\mu_0}(\rho)\big)\, d\rho \\
&\le \int_{r_1}^{r_0} \left[(H_0 + \rho_{\max} H_1)(\rho - r_1) + \rho_{\max} H_1(\mu_0 - \mu_1)\right] d\rho \\
&= \frac{H_0 + \rho_{\max} H_1}{2}\delta^2 + \rho_{\max} H_1 \Delta\delta \\
&\le \left(\frac{1}{2} H_0 + \frac{3}{2}\rho_{\max} H_1\right)\Delta^2.
\end{aligned}$$

The second case is $r_1 > r_0$. Then $\mu_1 > \mu_0$. If $r_0 = \rho_{\max}$, monotonicity would force $r_1 = r_0$, so here r_0 is an interior maximizer and $q_{\mu_0}(r_0) = 1$. Therefore

$$\begin{aligned}
\Gamma(r_0, \mu_0) - \Gamma(r_1, \mu_0) &= \int_{r_0}^{r_1} S(\rho - \mu_0)\big(q_{\mu_0}(\rho) - 1\big)\, d\rho \\
&\le \int_{r_0}^{r_1} (H_0 + \rho_{\max} H_1)(\rho - r_0)\, d\rho \\
&= \frac{H_0 + \rho_{\max} H_1}{2}\delta^2 \le M_\Gamma \Delta^2.
\end{aligned}$$

Thus (22) holds in all cases. Taking $\mu_0 = x^\top \theta^*$ and $\mu_1 = x^\top \theta$ gives

$$\Gamma(\rho^*(x^\top\theta^*), x^\top\theta^*) - \Gamma(\rho^*(x^\top\theta), x^\top\theta^*) \le M_\Gamma\big(x^\top(\theta - \theta^*)\big)^2,$$

which is exactly Lemma 3, with the coefficient in (19).

B OCO Properties of Link Adaptation Feedback (Lemma 4)

Proof. Fix a realized tuple (x_t, ρ_t, a_t) and a parameter $\theta \in \mathcal{H}$. Let

$$\omega := \rho_t - x_t^\top \theta \in \Omega, \qquad S(\omega) = 1 - F(\omega).$$

For an ACK, $a_t = 1$, the loss is $-\log S(\omega)$; for a NACK, $a_t = 0$, the loss is $-\log F(\omega)$. Direct differentiation gives

$$\nabla \ell_t(\theta) = \left[-a_t \frac{f(\omega)}{S(\omega)} + (1 - a_t)\frac{f(\omega)}{F(\omega)}\right] x_t, \tag{23}$$

$$\nabla^2 \ell_t(\theta) = \left[a_t \frac{f'(\omega)S(\omega) + f(\omega)^2}{S(\omega)^2} + (1 - a_t)\frac{f(\omega)^2 - F(\omega)f'(\omega)}{F(\omega)^2}\right] x_t x_t^\top. \tag{24}$$

The two scalar coefficients in (24) are exactly $-\frac{d^2}{d\omega^2}\log(1-F(\omega))$ and $-\frac{d^2}{d\omega^2}\log F(\omega)$, respectively. By the definition of κ_c in (3),

$$\nabla^2 \ell_t(\theta) \succeq \kappa_c x_t x_t^\top. \tag{25}$$

Similarly, (23) and the definition of ν_e in (4) imply

$$\nabla \ell_t(\theta) \nabla \ell_t(\theta)^\top \preceq \nu_e x_t x_t^\top. \tag{26}$$

Combining (25) and (26) yields

$$\nabla^2 \ell_t(\theta) \succeq \kappa_c x_t x_t^\top \succeq \frac{\kappa_c}{\nu_e} \nabla \ell_t(\theta) \nabla \ell_t(\theta)^\top.$$

Thus each realized loss is convex and α-exp-concave on $\mathcal{H}$ with the explicit parameter

$$\alpha := \frac{\kappa_c}{\nu_e}. \tag{27}$$

The gradient norm is also explicitly bounded:

$$\|\nabla \ell_t(\theta)\|_2^2 \le \nu_e \|x_t\|_2^2 \le \nu_e R_x^2. \tag{28}$$

All dependencies are through R_x, the compact interval Ω determined by $R_x, R_\theta, \rho_{\max}$, and the calibrated CDF F. □

C Proof of the Main Regret Theorem (Theorem 2)

Proof. The proof is pathwise up to the final conditional-expectation step. Define the parameter diameter and gradient bound

$$D_\theta := \sup_{u,v \in \mathcal{H}} \|u - v\|_2 \le 2R_\theta, \qquad G_\ell := R_x \sqrt{\nu_e}. \tag{29}$$

Let $\alpha = \kappa_c/\nu_e$ as in (27). If $D_\theta G_\ell = 0$, then either the parameter set is a singleton in all observed directions or the likelihood gradients vanish, and the displayed bound below is trivial. Otherwise set

$$\beta := \frac{1}{2}\min\left\{\frac{\kappa_c}{\nu_e}, \frac{1}{4 D_\theta G_\ell}\right\} \ge \frac{1}{2}\min\left\{\frac{\kappa_c}{\nu_e}, \frac{1}{8 R_\theta R_x \sqrt{\nu_e}}\right\}. \tag{30}$$

We run the undiscounted ONS update with $A_0 = \lambda_0 I_d$, $\lambda_0 > 0$, and a standard ONS step size satisfying

$$\eta \ge \frac{1}{\beta}. \tag{31}$$

The common canonical choice is $\eta = 1/\beta$.

Step 1: an explicit ONS likelihood-regret bound. For any comparator $u \in \mathcal{H}$, the exp-concavity inequality gives, for every realized loss,

$$\ell_t(\theta_t) - \ell_t(u) \le g_t^\top(\theta_t - u) - \frac{\beta}{2}\big(g_t^\top(\theta_t - u)\big)^2, \qquad g_t := \nabla \ell_t(\theta_t). \tag{32}$$

Indeed, $2\beta \le \alpha$, so Lemma 4 also gives 2β-exp-concavity. Setting $z = 2\beta g_t^\top(u - \theta_t)$ gives $|z| \le 1/4$ by (30). The inequality $-\log(1-z) \ge z + z^2/4$ on this interval then yields (32).

The A_t-projection step with $A_t = A_{t-1} + g_t g_t^\top$ and $\theta_{t+1} = \Pi_{\mathcal{H}}^{A_t}(\theta_t - \eta A_t^{-1} g_t)$ gives the following one-step inequality. Let $y_t = \theta_t - \eta A_t^{-1} g_t$. By the definition of the A_t-norm projection,

$$\|\theta_{t+1} - u\|_{A_t}^2 \le \|y_t - u\|_{A_t}^2.$$

Expanding the right-hand side gives

$$\|y_t - u\|_{A_t}^2 = \|\theta_t - u\|_{A_t}^2 - 2\eta g_t^\top(\theta_t - u) + \eta^2 g_t^\top A_t^{-1} g_t.$$

Rearranging yields

$$g_t^\top(\theta_t - u) \le \frac{1}{2\eta}\left(\|\theta_t - u\|_{A_t}^2 - \|\theta_{t+1} - u\|_{A_t}^2\right) + \frac{\eta}{2} g_t^\top A_t^{-1} g_t. \tag{33}$$

Since

$$\|\theta_t - u\|_{A_t}^2 = \|\theta_t - u\|_{A_{t-1}}^2 + \big(g_t^\top(\theta_t - u)\big)^2,$$

combining (32) and (33) yields

$$\ell_t(\theta_t) - \ell_t(u) \le \frac{1}{2\eta}\left(\|\theta_t - u\|_{A_{t-1}}^2 - \|\theta_{t+1} - u\|_{A_t}^2\right) + \left(\frac{1}{2\eta} - \frac{\beta}{2}\right)\big(g_t^\top(\theta_t - u)\big)^2 + \frac{\eta}{2} g_t^\top A_t^{-1} g_t.$$

The middle term is nonpositive by (31). Summing over $t = 1, \dots, T$ gives

$$\sum_{t=1}^{T}\big(\ell_t(\theta_t) - \ell_t(u)\big) \le \frac{\lambda_0 D_\theta^2}{2\eta} + \frac{\eta}{2}\sum_{t=1}^{T} g_t^\top A_t^{-1} g_t. \tag{34}$$

By the matrix determinant lemma,

$$\det(A_t) = \det(A_{t-1})\big(1 + g_t^\top A_{t-1}^{-1} g_t\big),$$

and $g_t^\top A_t^{-1} g_t = r_t/(1 + r_t)$ with $r_t = g_t^\top A_{t-1}^{-1} g_t$. Since $r/(1+r) \le \log(1+r)$ for $r \ge 0$,

$$\sum_{t=1}^{T} g_t^\top A_t^{-1} g_t \le \log\frac{\det(A_T)}{\det(A_0)}.$$

Using $\mathrm{tr}(A_T) \le d\lambda_0 + TG_\ell^2$ and the arithmetic–geometric mean inequality for eigenvalues,

$$\log\frac{\det(A_T)}{\det(A_0)} \le d\log\left(1 + \frac{TG_\ell^2}{d\lambda_0}\right) = d\log\left(1 + \frac{TR_x^2\nu_e}{d\lambda_0}\right). \tag{35}$$

Substituting (35) into (34), we have the explicit pathwise likelihood-regret bound

$$\sum_{t=1}^{T}\big(\ell_t(\theta_t) - \ell_t(u)\big) \le \frac{\lambda_0 D_\theta^2}{2\eta} + \frac{\eta d}{2}\log\left(1 + \frac{TR_x^2\nu_e}{d\lambda_0}\right) \le \frac{2\lambda_0 R_\theta^2}{\eta} + \frac{\eta d}{2}\log\left(1 + \frac{TR_x^2\nu_e}{d\lambda_0}\right). \tag{36}$$

Step 2: population likelihood controls throughput regret. Let $\mathcal{F}_{t-1}$ denote the history before a_t is observed, including the current context x_t. The quantities θ_t and $\rho_t = \rho^*(x_t^\top\theta_t)$ are $\mathcal{F}_{t-1}$-measurable. Define the conditional population loss

$$L_t(\theta) := \mathbb{E}\left[\ell_t(\theta) \mid \mathcal{F}_{t-1}, x_t, \rho_t\right].$$

At the true parameter θ^*, the ACK probability is $S(\rho_t - x_t^\top\theta^*)$ and the NACK probability is $F(\rho_t - x_t^\top\theta^*)$. Using (23), these probabilities cancel, so $\nabla L_t(\theta^*) = 0$. Since each realized loss has Hessian at least $\kappa_c x_t x_t^\top$ by Lemma 4, the conditional population loss has the same curvature lower bound. Therefore

$$L_t(\theta_t) - L_t(\theta^*) \ge \frac{\kappa_c}{2}\big(x_t^\top(\theta_t - \theta^*)\big)^2. \tag{37}$$

Combining (37) with Lemma 3, whose coefficient is explicitly given in (19), gives the per-round inequality

$$\Gamma(\rho^*(x_t^\top\theta^*), x_t^\top\theta^*) - \Gamma(\rho^*(x_t^\top\theta_t), x_t^\top\theta^*) \le \frac{2M_\Gamma}{\kappa_c}\left(L_t(\theta_t) - L_t(\theta^*)\right). \tag{38}$$

Taking expectations and using the tower property,

$$\mathbb{E}\big[L_t(\theta_t) - L_t(\theta^*)\big] = \mathbb{E}\big[\ell_t(\theta_t) - \ell_t(\theta^*)\big].$$

Thus, with $u = \theta^*$ in (36),

$$\mathbb{E}[\mathrm{Reg}_T(\mathrm{SOLA})] \le \frac{2M_\Gamma}{\kappa_c}\left[\frac{\lambda_0 D_\theta^2}{2\eta} + \frac{\eta d}{2}\log\left(1 + \frac{TR_x^2\nu_e}{d\lambda_0}\right)\right] \tag{39}$$

$$\le \frac{2M_\Gamma}{\kappa_c}\left[\frac{2\lambda_0 R_\theta^2}{\eta} + \frac{\eta d}{2}\log\left(1 + \frac{TR_x^2\nu_e}{d\lambda_0}\right)\right]. \tag{40}$$

Expanding M_Γ from (19), the coefficient multiplying the bracket in (40) is

$$\frac{2M_\Gamma}{\kappa_c} = \frac{1}{\kappa_c}\left[\sup_{\omega\in\Omega}\frac{f(\omega)}{1-F(\omega)} + 3\rho_{\max}\sup_{\omega\in\Omega}\frac{f'(\omega)(1-F(\omega))+f(\omega)^2}{(1-F(\omega))^2}\right]. \tag{41}$$

Equations (40)–(41) give the promised non-asymptotic coefficient dependence on $R_x, R_\theta, \rho_{\max}, \lambda_0, \eta, \kappa_c, \nu_e$, and the calibrated CDF F restricted to the compact interval Ω.

Finally, since for every $a > 0$ and $z \geq 0$,

$$\log(1+az) \leq \max\{1, a\}\log(1+z),$$

(40) implies the displayed main-text form with

$$C_0 = \frac{2M_\Gamma}{\kappa_c}\cdot\frac{2\lambda_0 R_\theta^2}{\eta},$$
$$C_1 = \frac{2M_\Gamma}{\kappa_c}\cdot\frac{\eta}{2}\max\left\{1, \frac{R_x^2\nu_e}{\lambda_0}\right\}.$$

No term in C_0 or C_1 depends on T. □

D Proof of the Online-Regression Lower-Bound Benchmark

As stated in the main text, this result is a related benchmark for the logarithmic statistical scale; it is not a direct minimax lower bound for the censored throughput problem.

Lemma 5 (Lower bound for online linear regression with squared loss)**.** *Let $d, T \geq 1$. In the online linear regression protocol, at each round t, the learner observes $x_t \in \mathbb{R}^d$, outputs a prediction $\hat{y}_t \in \mathbb{R}$, then observes $y_t \in \mathbb{R}$ and incurs loss*

$$(\hat{y}_t - y_t)^2.$$

Define the regret with respect to the best linear predictor in hindsight by

$$R_T = \sum_{t=1}^{T}(\hat{y}_t - y_t)^2 - \inf_{u\in\mathbb{R}^d}\sum_{t=1}^{T}\big(y_t - \langle u, x_t\rangle\big)^2.$$

Then

$$\inf_A \sup_{\substack{x_1,\dots,x_T\in\mathbb{R}^d \\ \|x_t\|_2\leq 1,\ y_1,\dots,y_T\in\mathbb{R}}} \mathbb{E}_A[R_T] \ \geq\ d\,H_{\lfloor T/d\rfloor},$$

where $H_m = \sum_{k=1}^{m}\frac{1}{k}$ is the m-th harmonic number (with the convention $H_0 = 0$). In particular,

$$\inf_A \sup_{\substack{x_1,\dots,x_T\in\mathbb{R}^d \\ \|x_t\|_2\leq 1,\ y_1,\dots,y_T\in\mathbb{R}}} \mathbb{E}_A[R_T] \ \geq\ d\log\big(\lfloor T/d\rfloor + 1\big).$$

Hence the minimax regret is

$$\Omega\big(d\log(1+T/d)\big),$$

and therefore $\Omega(d\log T)$ whenever d is fixed.

Proof. Fix an arbitrary online forecaster A. We first prove the claim for deterministic forecasters. The randomized case then follows by conditioning on the forecaster's internal random seed and averaging.

Set

$$m = \lfloor T/d\rfloor.$$

We construct a random sequence of examples as follows.

Let $\theta = (\theta_1, \dots, \theta_d)$ have independent coordinates with law $N(0,1)$. Let $(\xi_t)_{t\geq 1}$ be i.i.d. $N(0,1)$, independent of θ.

For rounds $t = 1, \dots, md$, define

$$j_t = 1 + ((t-1) \bmod d), \qquad x_t = e_{j_t},$$

where $e_1, \dots, e_d$ are the canonical basis vectors of $\mathbb{R}^d$, and set

$$y_t = \langle \theta, x_t \rangle + \xi_t = \theta_{j_t} + \xi_t.$$

For the remaining rounds $t = md+1, \dots, T$, set

$$x_t = 0, \qquad y_t = 0.$$

Clearly $\|x_t\|_2 \le 1$ for every t.

We will show that, under this random construction,

$$\mathbb{E}[R_T] \ge dH_m.$$

Since the expectation of a random variable is bounded above by its supremum over realizations, this implies that for the fixed forecaster A there exists at least one deterministic realization of the sequence $(x_t, y_t)_{t=1}^T$ such that

$$R_T \ge dH_m.$$

Taking the supremum over deterministic sequences then gives the desired lower bound for this forecaster. Since A was arbitrary, the minimax lower bound follows.

It remains to prove

$$\mathbb{E}[R_T] \ge dH_m.$$

Since the comparator $u = \theta$ is feasible, we have

$$R_T \ge \sum_{t=1}^{T} \Big((\hat{y}_t - y_t)^2 - (\langle \theta, x_t \rangle - y_t)^2 \Big).$$

For $t > md$, we have $x_t = 0$ and $y_t = 0$, so these rounds contribute a nonnegative quantity. Hence

$$R_T \ge \sum_{t=1}^{md} \Big((\hat{y}_t - y_t)^2 - (\theta_{j_t} - y_t)^2 \Big).$$

Let $\mathcal{F}_{t-1}$ be the σ-field generated by

$$(x_1, y_1, \hat{y}_1, \dots, x_{t-1}, y_{t-1}, \hat{y}_{t-1}, x_t).$$

Because the forecaster is deterministic, $\hat{y}_t$ is $\mathcal{F}_{t-1}$-measurable. Also, $y_t = \theta_{j_t} + \xi_t$, where ξ_t is independent of $\mathcal{F}_{t-1}$ and has mean zero. Therefore

$$\begin{aligned} &\mathbb{E}\big[(\hat{y}_t - y_t)^2 - (\theta_{j_t} - y_t)^2 \mid \mathcal{F}_{t-1}\big] \\ &\quad = \mathbb{E}\big[(\hat{y}_t - \theta_{j_t} - \xi_t)^2 - \xi_t^2 \mid \mathcal{F}_{t-1}\big] \\ &\quad = \mathbb{E}\big[(\hat{y}_t - \theta_{j_t})^2 \mid \mathcal{F}_{t-1}\big]. \end{aligned}$$

Now decompose around the conditional mean:

$$\begin{aligned} \mathbb{E}\big[(\hat{y}_t - \theta_{j_t})^2 \mid \mathcal{F}_{t-1}\big] &= \Big(\hat{y}_t - \mathbb{E}[\theta_{j_t} \mid \mathcal{F}_{t-1}]\Big)^2 + \mathrm{Var}(\theta_{j_t} \mid \mathcal{F}_{t-1}) \\ &\ge \mathrm{Var}(\theta_{j_t} \mid \mathcal{F}_{t-1}). \end{aligned}$$

Consequently,

$$\mathbb{E}[R_T] \ge \sum_{t=1}^{md} \mathbb{E}[\mathrm{Var}(\theta_{j_t} \mid \mathcal{F}_{t-1})].$$

Thus the problem reduces to computing the posterior variance.

Fix a coordinate $j \in \{1, \dots, d\}$. Whenever $x_t = e_j$, the learner observes one noisy sample of the form

$$Y = \theta_j + \xi, \qquad \xi \sim N(0,1).$$

The prior on θ_j is $N(0,1)$. Therefore, after k previous observations of coordinate j, the posterior distribution of θ_j is Gaussian with precision $1+k$, hence variance

$$\mathrm{Var}(\theta_j \mid \mathcal{F}_{t-1}) = \frac{1}{k+1}.$$

Equivalently,

$$\theta_j \mid \mathcal{F}_{t-1} \sim N\left(\frac{k}{k+1}\,\bar{Y}_{j,k}, \frac{1}{k+1}\right),$$

where $\bar{Y}_{j,k}$ is the average of the k previous observations from coordinate j.

In our construction, among the first md rounds each coordinate is visited exactly m times, and for each coordinate the numbers of previous observations are precisely

$$k = 0, 1, \ldots, m-1.$$

Therefore,

$$\mathbb{E}[R_T] \geq \sum_{j=1}^{d} \sum_{k=0}^{m-1} \frac{1}{k+1} = d \sum_{k=1}^{m} \frac{1}{k} = dH_m.$$

Finally, by the standard integral lower bound for harmonic numbers,

$$H_m \geq \int_1^{m+1} \frac{dx}{x} = \log(m+1).$$

Hence

$$\mathbb{E}[R_T] \geq d\log(m+1) = d\log\big(\lfloor T/d \rfloor + 1\big).$$

This completes the proof. □

Remark 6. The exact dependence proved above is

$$\Omega\big(d\log(1+T/d)\big).$$

Thus the common shorthand $\Omega(d\log T)$ is correct when d is treated as fixed. If one additionally imposes bounded responses and a bounded comparator class, one can prove analogous lower bounds by related Bayesian or information-theoretic arguments, but the Gaussian-prior proof above is the cleanest route in the unconstrained comparator setting.

E Additional Related Work

Dynamic pricing and binary purchase feedback. There is a mathematical analogy between the threshold-throughput abstraction used here and online posted-price models: the selected rate ρ_t plays a role similar to a posted price, the ACK/NACK outcome is analogous to a binary purchase decision, and goodput $\rho_t a_t$ is analogous to revenue from a successful sale. Classical online pricing work studies regret from posted prices when demand must be learned online (Kleinberg & Leighton, 2003). Feature-based dynamic pricing further incorporates contextual product or customer features into the pricing decision (Cohen et al., 2020). The closest pricing analogue to our known-CDF likelihood model is Javanmard & Nazerzadeh (2019), who study high-dimensional contextual pricing with binary-valued sales feedback and a known market-noise distribution.

The distinction from the known-noise RMLP guarantee is in the online regime: SOLA uses per-slot second-order updates under arbitrary bounded non-anticipating contexts, rather than episodic sparse estimation under i.i.d. contexts with a nondegenerate second moment. The bounded-rate analysis explicitly treats upper-boundary optima. These differences do not establish novelty for the generic threshold-feedback/ONS template or a sparsity-adaptive improvement. The wireless development connects ACK/NACK likelihood curvature to delivered-goodput regret and motivates separate empirical tests of finite MCS, target BLER, calibration, and forgetting.

F Experiment Details and Reproducibility Notes

Table 1 records the original synthetic settings, fixed across seeds and noise families. All SOLA/GLM-UCB variants use Euclidean projection onto the radius-5 ball and regularization 1. Rate selection and the oracle use the 80-point grid $0.05, 0.10, \ldots, 4.0$. These practical settings do not verify the continuous-rate theorem's numerical conditions. OLLA instead takes continuous actions in $[0.01, 6]$, so its expected-throughput gap to this grid oracle need not be nonnegative.

Table 1: Synthetic experiment settings.

Setting	Stationary synthetic experiment	Abrupt-change stress test
Horizon and runs	$T = 20{,}000$, $d = 10$; ten runs per regime	$T = 30{,}000$, $d = 10$; ten seeds $42, \ldots, 51$
Contexts	Bias-augmented normalized contexts with $\|x_t\|_2 = 1$	Same context generator
Parameter model	Fixed θ^*; bias 1.0, non-bias component norm 1.5	At 10,000, flip non-bias signs; at 20,000, redraw θ^*
Noise	Gaussian $\sigma = 0.30$; Gumbel scale 0.234 with centered location; logistic scale 0.165	Gaussian $\sigma = 0.30$
SOLA / D-SOLA	ONS multiplier 1; no discount or margin	ONS multiplier 1; discount 0.998; no margin
GLM-UCB / D-GLM-UCB	$\beta = 1$, update multiplier 1; no discount	$\beta = 0.1$, update multiplier 0.5; discount 0.995
OLLA	Initial rate 1; ACK step $+0.1$, NACK step -0.9; target BLER 0.1	Same steps and target; initial rate 2 in the runner

Original curve conventions. Figure 1 plots only the ten-run mean, with $t \geq 10$ and a 10^{-3} plotting floor. The original time-series arrays are unavailable, so we cannot check which values, if any, were affected by that floor. Figure 2 records parameter error before updating and expected regret against each segment's grid oracle; its SD bands use divisor ten (`ddof=0`).

Trace retention and environment construction. The audit covers 83 CSV sessions with CQI/SNR/RSRP/DL_bitrate fields. Four have fewer than 500 rows; two fail the nondegeneracy check below. The 77 retained horizons are $T = \min\{\text{rows}, 20{,}000\}$, ranging from 615 to 7,190. After numeric conversion, missing values are forward-filled, backward-filled, then set to zero if still missing. The three radio features are centered and divided by their whole-session population SD plus 10^{-8}; adding a bias gives $d = 4$, without row normalization.

DL_bitrate is smoothed by a trailing ten-row mean (minimum one row). Dividing by its session 95th percentile, multiplying by 3.5, and clipping to $[0.05, 4]$ gives a response proxy; a percentile at most 10^{-5} is replaced by 1. A ridge fit with penalty 10^{-6}, including the intercept, predicts this proxy from the contexts. Let $q_{.05}, q_{.95}$ be the fitted prediction quantiles. Sessions with $q_{.95} - q_{.05} < 10^{-8}$ are excluded; otherwise an affine rescaling maps these quantiles to 0.5 and 3.5. This yields $\mu_t = x_t^\top \theta^*$ with a fixed per-session θ^*, without clipping μ_t. Normalization and fitting use the entire session: this constructs a controlled environment offline, rather than evaluating online prediction on held-out data. The bitrate proxy does not directly determine ACKs.

Trace feedback and methods. Capacity is $\mu_t + \epsilon_t$, where logistic noise has location zero and scale $0.30\sqrt{3}/\pi$; ACK is $\mathbb{I}\{\rho_t \leq \mu_t + \epsilon_t\}$. D-SOLA uses this matched CDF, discount 0.96, regularization 0.99, ONS multiplier 1, and Euclidean radius-5 projection. It maximizes predicted goodput on the 80-point $[0.05, 4]$ grid, subtracts a global margin in $\{0, 0.25, 0.50, 0.75, 1\}$, then clips to $[0.05, 4]$ without re-quantizing. Trace OLLA is the project's scalar implementation: initial rate 1, ACK step $+0.0016$, NACK step -0.0064, and the same bounds (target BLER 0.20).

SALAD configuration and stationary comparison. The adapter calls the unmodified NVlabs SALAD implementation (commit `2746947`) and its Student SINR estimator, using constructor defaults without knowledge distillation. It supplies the rate table and target BLER explicitly: 80 rates/0.20 for traces, 30 rates/0.10 for the stationary comparison. Logistic BLER curves match the simulated channel. Batch size is one; each ACK/NACK selects the next slot's MCS, without additional HARQ delay. The stationary experiment uses $T = 5{,}000$, $d = 1$, $x_t = 1$, $\mu_t = 2.2$, logistic noise SD 0.30, and 30 equally spaced rates in $[0.05, 4]$. Its OLLA comparator uses the official SALAD repository's estimator, unlike trace OLLA, and obtains goodput 1.585 ± 0.008 and BLER 0.1002 ± 0.0001 (sample SD across ten paired seeds).

Aggregation and CDF comparisons. Trace metrics first average ten shared-noise seeds per session, then weight the 77 sessions equally, regardless of horizon. Figure 3 bars are sample SDs of session

means (`ddof=1`). No margin is tuned by session or seed; 0.25 is highlighted descriptively. Paired CIs use 4,000 whole-session bootstrap resamples and percentile endpoints, without selection adjustment. The CDF study uses $T = 20{,}000$, $d = 10$, ten paired seeds, common environmental randomness, and a true-CDF grid oracle. Its 22 conditions comprise scale multipliers $\{0.5, 0.75, 1, 1.25, 1.5, 2\}$ for each of Gaussian, logistic, and Gumbel noise, plus Gumbel/logistic and Gaussian/logistic swaps in both directions, matched in mean and variance. No condition is retuned. CDF CIs instead resample the ten paired synthetic runs 4,000 times. Both bootstrap procedures use RNG seed 1701 and 2.5/97.5 percentiles. The original synthetic runs used at most four CPU cores and under one hour per experiment; this budget does not cover all additional studies.

G More Discussions

This section complements Section 7.

A restart benchmark for forgetting. Suppose there are K stationary segments of lengths T_k, with known boundaries, a common known F, parameters in the same $\mathcal{H}$, and uniform bounds as in Theorem 2. Restarting the exact update of Algorithm 1, including its metric projection and continuous rate maximization, at every boundary gives

$$\mathbb{E}[\mathrm{Reg}_T(\text{restarted SOLA})] \le KC_0 + C_1 d \sum_{k=1}^{K} \log(1 + T_k/d).$$

Here regret is against each segment's parameter-informed oracle. The inequality follows by applying Theorem 2 conditionally on the history at each restart and summing. It motivates discarding stale information; it is not a guarantee for unknown change points or the implemented D-SOLA update.

Computational considerations. The Python microbenchmark sums component timings for rate selection, likelihood/gradient calculation, inverse-metric update, preconditioned step, and Euclidean projection; it excludes input generation and outer-loop overhead. On Darwin/arm64 with Python 3.12.7 and NumPy 1.26.4, BLAS thread settings were set to one. Each of five repetitions uses 50 warm-up and 300 measured slots with pregenerated contexts and binary ACKs. Pooling 1,500 slots per dimension gives medians $19.87, 45.46, 883.08\,\mu$s at $d = 10, 100, 500$, and p95 values $22.50, 51.25, 1018.06\,\mu$s. The fitted exponent for $d \ge 50$ is 1.90. These component sums do not measure end-to-end scheduler latency or arbitrary metric projections; the exact CPU model and actual BLAS thread count were not recorded.

Future work and broader impact. Open directions include guarantees under parameter variation, reliability constraints, finite MCS, and CDF misspecification, as well as link-level and testbed validation. Better adaptation may reduce wasted spectrum and retransmissions, but aggressive rate selection can increase outage or interact poorly with scheduler load. Deployment would require BLER monitoring, checks across radio conditions, and rollback mechanisms beyond the statistical model studied here.