When Loss Signals Dominate Context: Adaptive Expert Routing in the Loss-Dominance Regime

Samridhi Saini* **Vatsal Khanna***
Dr. B. R. Ambedkar National Institute of Technology Jalandhar
{samridhis.ec.23, vatsalk.ec.23}@nitj.ac.in

Abstract

Modern adaptive systems increasingly rely on learned contextual models, often deep networks to anticipate regime changes and guide decision-making. Intuitively, such context should improve performance by accelerating the identification of the best expert. However, this intuition implicitly assumes that learned signals provide information faster or more reliably than the system's own feedback. In this work, we show that this assumption fails in a natural and identifiable class of settings. We study adaptive expert routing under full-information feedback and identify a regime in which online loss signals alone are sufficient for near-optimal performance. When inter-expert performance gaps are large, loss-based routing rapidly concentrates on the optimal expert, leaving little room for improvement from contextual predictions. In this regime, even moderately accurate predictors introduce persistent bias that outweighs their potential benefit. We formalize this phenomenon through a mixing-time analysis and a regret decomposition, leading to a computable condition under which contextual information does not improve routing performance in this regime. This characterization yields a practical diagnostic based on context-weight interpolation that determines whether a given system operates in this loss-sufficiency regime. Empirical evaluation in a real-time, microsecond-latency adaptive control system (16 kHz Active Noise Control) confirms the theory: loss-based routing matches oracle performance, while contextual routing does not provide measurable improvement and often degrades performance in our setting.

1 Introduction

Many engineering and machine learning systems operate a bank of adaptive algorithms in parallel, each embodying a different inductive bias or optimization strategy. Such settings arise in adaptive control, signal processing, and online model selection, where no single algorithm performs well across all regimes. A central challenge is therefore how to combine these experts in real time: which expert to trust, and whether this decision should incorporate additional contextual information.

A common approach augments online routing with learned context, using predictive models to infer the current regime and guide expert selection. The intuition is that accurate context should accelerate identification of the best expert. This implicitly assumes that learned context provides useful information faster or more reliably than the system's own feedback.

This paper examines the validity of this assumption and identifies a regime in which it fails. We consider adaptive expert routing under full-information feedback, where all expert losses are observed at each step, providing a high-frequency signal reflecting expert quality. The key question is whether

*Equal contribution.

ICML 2026 Workshop on Decision-Making from Offline Datasets to Online Adaptation: Black-Box Optimization to Reinforcement Learning

contextual information adds value beyond what this signal already resolves. Our focus is on regimes with large inter-expert performance gaps and fast loss-based identification.

Central Question. *When does online loss feedback alone suffice for adaptive expert routing, and when can additional contextual information provide a genuine advantage?*

We formalize this setting and analyze loss-based routing dynamics. Our analysis yields bounds on the rate of optimal expert identification and quantifies the effect of imperfect contextual predictions, leading to a computable condition under which loss signals are sufficient and contextual information cannot improve performance-an effect driven by the relative information rate of feedback versus context, not by limitations of any particular predictive model. Understanding when this occurs is important for designing adaptive systems, including online model selection and related settings where feedback is rapidly available.

Contributions.

- **Problem formulation.** We formalize adaptive expert routing under full-information feedback with heterogeneous, simultaneously adapting experts, distinguishing it from mixture-of-experts, contextual bandits, and algorithm selection.
- **Mixing-time characterization.** We derive bounds on the time required for loss-based routing to identify the optimal expert as a function of inter-expert performance gaps.
- **Context-sufficiency condition.** We show that when loss signals resolve expert identity sufficiently quickly, incorporating imperfect contextual predictions increases expected regret, yielding a computable condition under which context cannot improve performance.
- **Practical diagnostic.** We introduce a simple context-weight interpolation diagnostic to determine whether a system operates in this regime.
- **Empirical validation.** We validate the analysis in a real-time adaptive control system (16 kHz Active Noise Control), a microsecond-latency setting that limits the usefulness of learned context. Loss-based routing matches oracle performance, while contextual routing does not provide benefit.

2 Related Work

Online learning and expert tracking. Combining multiple experts has been widely studied in online learning, notably via Weighted Majority and Hedge [23, 11, 17, 34] and extensions to non-stationary environments [14, 5, 13, 29, 7]. These methods provide regret guarantees under loss sequences typically analyzed as independent of the learner's internal state [3, 36]. In contrast, our setting involves adaptive experts whose losses evolve with their internal state and the routing policy, introducing endogenous coupling between experts.

Mixture of Experts. Mixture of Experts (MoE) models [15, 16] combine predictors through a learned gating function, with modern variants using sparse routing [30, 22, 9, 37]. In such architectures experts are static and routing is learned offline, making routing a representation problem. Our setting instead involves continuously adapting experts with endogenous losses, where routing operates online and directly influences expert dynamics-a control problem, not a representation problem.

Contextual bandits and algorithm selection. Contextual bandits [20, 1, 6, 10] use side information to guide decisions under partial feedback [3, 21, 36], while algorithm selection [28, 18] leverages problem features to choose among algorithms, typically in offline or episodic settings. Our setting differs in assuming full-information feedback [21], where losses of all experts are observed, and in focusing on continuously adapting experts rather than static performance profiles.

Predictable sequences and auxiliary information. Incorporating auxiliary or predictable structure into online learning can improve performance when such information is sufficiently accurate [27]. These approaches assume auxiliary signals capture useful structure in the loss sequence. Our work instead examines how such signals interact with feedback that continuously reflects expert performance in adaptive settings.

Active noise control and adaptive filtering. Active noise control (ANC) is a classical signal processing problem [26, 19, 8], where adaptive filters such as FxNLMS [35, 25] and RLS variants are used for real-time disturbance cancellation. Prior work has explored combining adaptive filters using fixed or adaptive mixing strategies [2, 31]. Here, ANC serves as a high-frequency testbed for adaptive expert routing under stringent latency constraints, while the formulation and analysis extend beyond this domain.

3 Problem Formulation

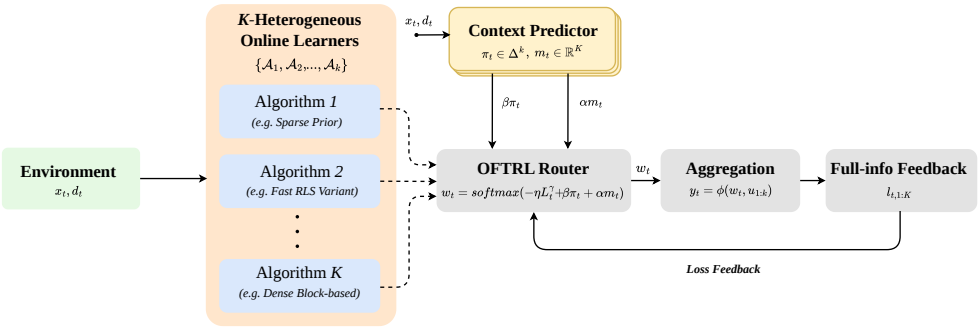

Figure 1: K=5 heterogeneous adaptive experts run in parallel; the OFTRL router concentrates weight on the best expert using discounted losses alone (Eq. 4), without any context. The context predictor is theoretically unnecessary and empirically harmful whenever loss-dominance holds (Definition 2, Theorem 3).

3.1 Expert Bank and Loss Feedback

Figure 1 gives an overview of the complete system. Let $[K] \triangleq \{1, \ldots, K\}$ be a bank of K adaptive algorithms. Each expert k maintains an internal state $\theta_{t,k}$, updated online, and maps a reference input x_t to an output $u_{t,k} = h_k(\theta_{t,k}, x_t)$, updating $\theta_{t,k}$ after each step. Given desired signal d_t, its loss is

$$\ell_t(k) \;=\; \big(d_t + S\, u_{t,k}\big)^2, \tag{1}$$

where S is the fixed linear *secondary-path* operator (acoustic transfer function from actuator to error sensor; more generally, any fixed map from expert output to observable error). The full loss vector $\ell_t = (\ell_t(1), \ldots, \ell_t(K)) \in \mathbb{R}^K$ is revealed to the router each round, establishing the *full-information* setting.

Endogenous coupling. Each expert updates $\theta_{t,k}$ from its own history, so $\ell_t(k)$ depends on past routing decisions $\{w_s\}_{s<t}$ through the state trajectory. Section 4 accounts for this by conditioning on the routing history; the theoretical results hold in the post-mixing regime where expert losses have stabilised near their stationary values (Appendix B).

3.2 Routing Policy and Objective

The router selects $w_t \in \Delta^K$ at each step, incurring aggregate loss

$$L_t = \langle w_t,\, \ell_t \rangle, \tag{2}$$

with objective being the minimization of the $\sum_{t=1}^{T} L_t$ relative to the best expert in hindsight, i.e. $k^* \triangleq \arg\min_k \sum_{t=1}^{T} \ell_t(k)$.

Optionally, the router may use a context vector c_t via a learned predictor $(m_t, \pi_t) = f(c_t)$ (architecture in Appendix C). Section 4 parameterises this through scalars $\alpha, \beta \geq 0$: $\alpha = \beta = 0$ is pure loss-driven routing; $\alpha > 0$ injects the loss hint; $\beta > 0$ injects the regime prior.

3.3 Routing Necessity

In a heterogeneous bank, each expert converges to a distinct *steady-state output* u_k^* under stationary inputs. Since these steady states optimise different trade-offs, they are generally incompatible [2], hence implying that uniform mixing does not recover any single expert's solution.

Proposition 1 (Routing Necessity)**.** *Let u_k^* be the steady-state output of expert k under squared-error loss $L(y,d) = (d + Sy)^2$. If $\exists\, j \neq k$ with $Su_j^* \neq Su_k^*$, then*

$$\mathbb{E}\big[L\big(\tfrac{1}{K}\textstyle\sum_k u_k^*,\ d\big)\big] \ > \ \min_k \ \mathbb{E}[L(u_k^*,\, d)]. \tag{3}$$

Averaging incompatible steady states introduces cross-terms $\mathbb{E}[(Su_j^* - Su_k^*)^2] > 0$ that strictly inflate the mixture loss (full derivation in Appendix A). The result holds for any algorithm family or step-size schedule; in our five-expert bank (Table 1) the condition is satisfied by design. Empirically, uniform routing yields -5.45 dB overall (Section 6), confirming that a missing router actively harms performance.

4 Theoretical Analysis

We characterise how quickly OFTRL concentrates on the optimal expert, quantify the marginal effect of injecting context predictions, and derive a computable condition under which $\beta = 0$ minimises expected regret.

4.1 Routing Algorithm

We consider discounted Follow-the-Regularised-Leader (OFTRL) with negative-entropy regularisation [13, 29, 24]:

$$w_t = \mathrm{softmax}\left(-\eta \sum_{s=1}^{t-1} \gamma^{t-1-s} \ell_s\right), \tag{4}$$

where $\eta > 0$ is the learning rate and $\gamma \in (0,1)$ down-weights stale losses. To study the role of context, we embed OFTRL in the two-parameter family

$$w_t^{\alpha,\beta} = \mathrm{softmax}\left(-\eta \sum_{s=1}^{t-1} \gamma^{t-1-s} \ell_s + \alpha\, m_t + \beta\, \pi_t\right), \quad \alpha, \beta \geq 0, \tag{5}$$

where $m_t \in \mathbb{R}^K$ is the short-horizon loss-hint vector and $\pi_t \in \Delta^K$ is the regime prior (both from the context predictor, Section 3.2); $\alpha = \beta = 0$ recovers OFTRL. The scalar $\beta \geq 0$ controls context injection via the regime prior; $\alpha \geq 0$ weights a loss-derived predictive hint. For theoretical analysis we study the α=0 reduction (denoted w_t^{β}), which isolates the cost of context injection. In the full deployed system, α is modulated by an OOD gate $g_t \in [0,1]$ derived from the OOD embedding head (d) (Appendix C): for in-distribution contexts $g_t \approx 1$; for out-of-distribution segments $g_t \to 0$, suppressing the loss-hint contribution. All theoretical results use g_t=1. Cumulative regret relative to k^* is

$$R_T(\beta) \ \triangleq \ \sum_{t=1}^{T} \langle w_t^{\beta},\, \ell_t \rangle - \sum_{t=1}^{T} \ell_t(k^*). \tag{6}$$

4.2 OFTRL Mixing Time

Definition 1 (Routing Mixing Time)**.** *For $\epsilon \in (0, \frac{1}{2})$,*

$$T_{\mathrm{mix}}(\epsilon) \ \triangleq \ \inf\big\{t \geq 1 : w_t(k^*) \geq 1 - \epsilon\big\}.$$

Let T_0 denote the end of the initial transient during which the loss gap may be small. Define the *minimum inter-expert gap*

$$\Delta_{\min} \ \triangleq \ \min_{k \neq k^*} \big[\ell_t(k) - \ell_t(k^*)\big] \ > \ 0, \quad t > T_0. \tag{7}$$

Theorem 1 (OFTRL Mixing Time)**.** *Under assumption* (7)*, for any* $\epsilon \in (0, \frac{1}{2})$*,*

$$T_{\text{mix}}(\epsilon) \;\le\; T_0 + \left\lceil \frac{\log\big(\frac{(K-1)(1-\epsilon)}{\epsilon}\big)}{\eta\,\Delta_{\min}} \right\rceil. \tag{8}$$

The bound is $O(\log K/(\eta\Delta_{\min}))$: larger performance gaps and larger learning rates yield faster concentration on k^*. The proof tracks the score gap in softmax logit space under discounted accumulation (Appendix A.2).

4.3 Context Injection and Its Cost

Let $p \triangleq \mathbb{P}[\arg\max_k \pi_{t,k} = k^*] < 1$ be the predictor's top-1 accuracy.

Theorem 2 (Context Injection Harm)**.** *For* $t > T_0 + T_{\text{mix}}(\epsilon)$*,*

$$\mathbb{E}\big[\langle w_t^\beta, \ell_t\rangle - \langle w_t^0, \ell_t\rangle\big] \;=\; -O(p\,\epsilon^2\beta) + O\big((1-p)\,\epsilon\,\beta\,\Delta_{\min}\big). \tag{9}$$

As $\epsilon \to 0$*, the right-hand side is positive under the stated assumptions and in the post-mixing regime.*

Remark 1. *This result relies on a local perturbation analysis around the post-mixing distribution and assumes approximate exogeneity of losses; extending the analysis to fully coupled adaptive dynamics remains open.*

After mixing, w_t^0 concentrates weight $1-\epsilon$ on k^*, so a correct prediction (probability p) can shift weight by at most $O(\epsilon^2)$, while an incorrect prediction (probability $1-p$) deflects weight away from k^* at $O(\epsilon)$, amplified by $\Delta_{\min}$. The gain is quadratic in ϵ; the harm is linear—an asymmetry that grows with the gap. Proof by decomposing expected loss into correct and incorrect prediction events (Appendix A.3).

Corollary 2.1 (Context Accuracy Threshold)**.** *Under the hypotheses of Theorem 2, context injection with* $\beta > 0$ *reduces expected per-step loss only when predictor accuracy satisfies*

$$p \;>\; p_{\textit{thresh}} \;\triangleq\; \frac{\Delta_{\min}}{\Delta_{\min} + \epsilon}. \tag{10}$$

For our system ($\Delta_{\min} = 25.9$*, the mean inter-expert gap across the 7 loss-dominated conditions marked* $\star$ *in Table 2;* $\epsilon = 0.05$*),* $p_{\textit{thresh}} \approx 0.998$*, far above the measured accuracy* $p = 0.761$*.*

4.4 Context-Sufficiency Condition

Let T_{window} be the *context window*: the minimum number of steps the predictor requires to form a confident prediction (e.g. the feature-extraction window length; 32 ms = 512 steps at 16 kHz in our system).

Theorem 3 (Context-Free Sufficiency)**.** *Suppose* $\Delta_{\min} > 0$*,* $p < 1$*, and* $K\,T_{\text{mix}}(\epsilon) \le T_{\textit{window}}$*. Then for all* $\beta > 0$ *and* $T \ge T_{\textit{window}}$*,*

$$\mathbb{E}[R_T(0)] \;\le\; \mathbb{E}[R_T(\beta)]. \tag{11}$$

The condition $KT_{\text{mix}} \le T_{\text{window}}$ formalises *loss-signal dominance*: OFTRL identifies k^* K times over before the predictor completes one confident observation. Under this condition, the transient regret of w^0 is outweighed by the accumulated post-mixing harm from Theorem 2, making $\beta = 0$ the optimal choice. The full regret decomposition and a practical guide to verifying the condition from oracle buffer data are in Appendix A.4.

Definition 2 (Loss-Dominance Regime)**.** *A routing system with context window* $T_{\textit{window}}$*, inter-expert gap* $\Delta_{\min}$*, and predictor accuracy* p *operates in the* loss-dominance regime *when*

$$K\,T_{\text{mix}}(\epsilon) \;\le\; T_{\textit{window}}, \quad \Delta_{\min} > 0, \quad p < 1, \tag{12}$$

where T_{mix} *is given by* (8)*. The online loss signal then resolves optimal-arm identity* K *times over before the predictor completes a single confident observation, making* $\beta{=}0$ *optimal in expectation (Theorem 3).*

5 System Design and Experimental Setup

We now instantiate and verify the three conditions of Definition 2 in a concrete real-time system, then report results of the full experimental protocol.

5.1 Expert Bank

We maintain K=5 adaptive experts $\{\mathcal{A}_k\}_{k=1}^{K}$ drawn from two algorithm families: filtered-x LMS [35, 25] and recursive least-squares variants [12]. Expert outputs $u_{t,k} = h_k(\theta_{t,k}, x_t)$, per-step losses $\ell_t(k) = (d_t + Su_{t,k})^2$, and secondary-path operator S are as defined in Section 3; all K losses are observed each step (full-information setting).

The expert bank is constructed to exhibit heterogeneous behaviour across operating conditions so that the identity of the optimal expert varies across regimes, making static or uniform aggregation suboptimal (Proposition 1). Outputs are clipped to $\pm 3\hat{\sigma}_d$ (three-sigma envelope of the disturbance estimate) to prevent transient amplification; this constraint is inactive under nominal operation. Full expert specifications are in Appendix B.

5.2 Context Predictor

At each time step t, a context vector c_t is constructed from recent observations. A multi-head predictor $f(c_t)$ produces four outputs; heads (a) and (b) enter the routing update (5) directly:

$$(m_t,\, \pi_t) = f(c_t), \quad m_t \in \mathbb{R}^K,\ \pi_t \in \Delta^K, \tag{13}$$

where m_t is a short-horizon loss-hint vector (head (a)) and π_t is a soft regime prior over experts (head (b)). Auxiliary heads (c) and (d) serve routing-support roles described in Appendix C: head (c) provides a per-expert multiplicative mask (the *expert mask*) applied after (5), and head (d) provides an OOD embedding used to gate the loss-hint coefficient α.

The predictor is a dual-stream Temporal Convolutional Network [4] trained via supervised distillation from oracle expert labels. The regime prior has top-1 accuracy $p = \mathbb{P}[\arg\max_k \pi_{t,k} = k^*] < 1$, consistent with the assumptions of Theorem 2. The scalars α and β in (5) gate heads (a) and (b) independently: α=0 disables the loss hint; β=0 disables the regime prior; α=β=0 reduces to pure OFTRL. The full system uses α=0.50, β=0.50. Architectural design, feature construction, and training details are in Appendix C.

5.3 Evaluation Protocol

Evaluation conditions. We evaluate over 13 acoustic regimes (6 stationary and 7 non-stationary), spanning tonal, broadband, impulsive, switching, and mixed noise classes. Each run consists of T = 8,000 steps (500 ms) with a 2,000-step burn-in period. Experiments are conducted over 5 random seeds, yielding 65 runs per method. Performance is measured via noise reduction

$$\mathrm{NR} = -10\log_{10}\left(\frac{\bar{\ell}_{\mathrm{eval}}}{\bar{\ell}_{\mathrm{init}}}\right)\ \mathrm{dB}, \tag{14}$$

where $\bar{\ell}_{\mathrm{eval}}$ denotes the mean loss after burn-in and $\bar{\ell}_{\mathrm{init}}$ the initial loss.

Baselines. We compare the following routing strategies: (i) *Oracle*, which selects the best expert in hindsight, (ii) *OFTRL* ($\alpha = \beta = 0$), (iii) *Hedge* with learning rates $\eta \in \{0.10, 0.01\}$, and (iv) *Uniform routing*, $w_t = K^{-1}\mathbf{1}$. All methods operate on the same expert bank; differences arise solely from the routing policy.

Composite score. When comparing routing policies across conditions with heterogeneous NR magnitudes, we use the composite score $\overline{\mathrm{NR}} - \frac{1}{2}\sigma_{\mathrm{NR}}$, where $\overline{\mathrm{NR}}$ is the mean NR across all 13 conditions and σ_{NR} its standard deviation. This penalises methods that achieve high mean NR at the cost of high cross-condition variance.

Statistical testing. Results are reported with 95% bootstrap confidence intervals (10,000 replicates), paired by seed and condition. Bootstrap estimation is used due to non-Gaussian performance distributions across regimes.

6 Results

6.1 Routing Necessity

Per-condition NR for all five routing policies is reported in Table 7 (Appendix E). Uniform routing achieves -5.45 dB overall (-12.1 dB on wideband-structured noise), confirming Proposition 1: destructive interference from incompatible steady-state filter solutions makes uniform mixing harmful across almost every condition.

OFTRL achieves $+3.94$ dB, within 0.22 dB of the hindsight oracle ($+4.16$ dB). On the two switching conditions OFTRL *exceeds* the oracle: $+4.02$ vs. $+2.13$ dB (switch_250ms) and $+2.14$ vs. $+1.48$ dB (switch_fast). The oracle incurs a hard-switch penalty at each regime change; OFTRL's soft mixture smoothly redistributes weight during transitions, eliminating this cost.

Hedge (η=0.01, $+2.88$ dB) and Hedge (η=0.10, $+2.94$ dB) lag OFTRL by ≈ 1.0 dB overall; the gap widens to 1.7–2.7 dB on tonal conditions where fast arm identification matters most. The advantage of discounting over equal-weight accumulation is consistent with the theory: EWA (Hedge with γ=1) cannot adapt to regime changes because past losses from different regimes contaminate the current posterior.

6.2 The β-Injection Diagnostic

The central empirical result is the composite score $(\bar{\mathrm{NR}} - \frac{1}{2}\sigma_{\mathrm{NR}})$ (defined in Section 5.3) as a function of context injection strength β across six evaluation horizons. Score is monotonically decreasing in β at every horizon, with β^*=0 universally. Adding any context signal at any evaluation length from 62.5 ms to 1 s strictly degrades routing quality.

Quantitative summary. At T=16,000 steps (1 s), the score drops from 5.78 at β=0 to 4.86 at β=5 (-16.0%). At T=1,000 steps (62.5 ms), the drop is from 1.10 to 0.83 (-23.9%). The *magnitude* of harm grows with T (consistent with Theorem 2: post-mix harm accumulates over $T - T_{\mathrm{mix}}$ steps) but is significant even at the shortest horizon.

Why harm occurs even at T=1,000. The diagnostic uses a *fixed* burn-in of 500 steps across all horizons (the main evaluation uses a proportional 2,000-step burn-in for T=8,000; both follow oracle-buffer warm-up protocol). In high-gap regimes ($\Delta_{\min} \approx 22$), $T_{\mathrm{mix}} \approx 20$ steps, so $w_{t,k^*} > 0.95$ by step 25 of the evaluation window-well before the predictor's 32-ms (512-step) feature window completes. For the $\approx 96\%$ of post-mixing steps the predictor introduces persistent bias; the asymmetry of Theorem 2 ensures harm dominates the small transient benefit. Figure 2 shows this relationship at six evaluation horizons.

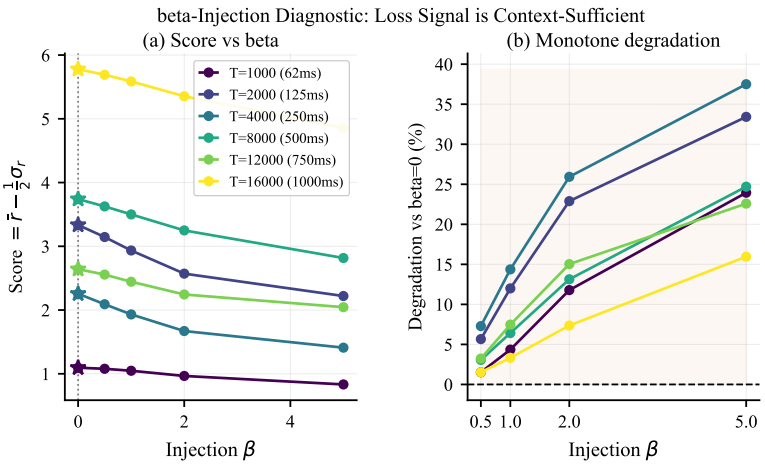

Figure 2: Composite score vs. β at six horizons (62.5 ms–1 s). Monotone decrease at every horizon; β^*=0 universally (Theorem 3).

6.3 Expert Bank Dominance Structure

The oracle buffer logs the full loss vector ℓ_t under a passive evaluation policy across 1.97M steps, enabling post-hoc computation of each expert's win-share (fraction of steps with strictly minimum loss) and the empirical dominance gap $\Delta_{\min}$. Tonal and structured conditions exhibit near-total dominance by one expert (win-share > 0.80, dominance gap > 0.40), corresponding to $\Delta_{\min} \approx 20$–50 and $T_{\text{mix}} \leq 20$ steps. These are precisely the conditions where OFTRL most strongly outperforms Hedge. Broadband and impulsive conditions show smaller gaps (win-share < 0.40); no expert dominates, T_{mix} is large, and no routing algorithm can substantially improve over the bank mean. This bimodal gap structure empirically grounds the three conditions of Theorem 3. Figure 3 visualises the win-share distribution across all 13 regimes.

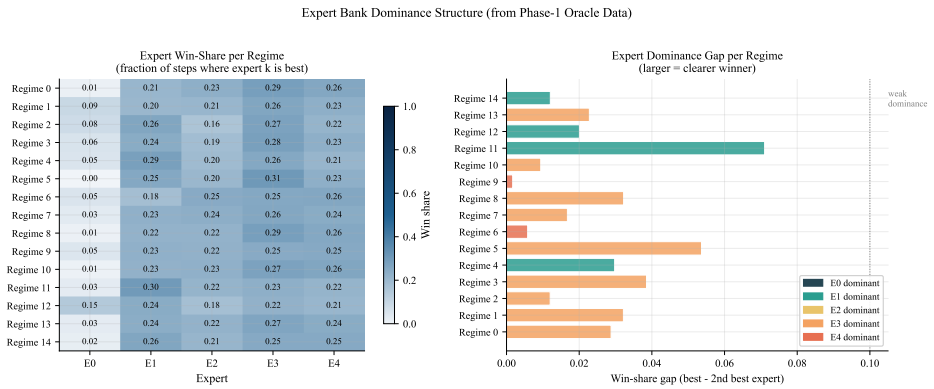

Figure 3: Win-share heatmap: K=5 experts $\times$ 13 regimes. Tonal conditions show near-total RLS dominance; broadband/impulsive show distributed shares with small $\Delta_{\min}$ (Table 2).

6.4 Ablation Study

Table 4 (Appendix D) reports a 5-way ablation over 325 cells. The `no_predictor` ablation ($\alpha = \beta = 0$) achieves $+3.86$ dB, exceeding the `full` system ($+3.49$ dB) with statistical significance (Δ=$+0.37$ dB, 95% CI $[+0.22, +0.52]$, p<0.001). Disabling the expert mask alone yields $+3.85$ dB (Δ=$+0.36$ dB, 95% CI $[+0.21, +0.51]$). In every ablation direction, removing predictor components uniformly improves or maintains performance.

Loss-only routing. The β=0 ablation achieves $+3.86$ dB, equal to `no_predictor` and exceeding the uncalibrated `full` system ($+3.49$ dB), confirming that Phase-3 calibration correctly recovers β^*=0 as the optimal injection strength.

6.5 Robustness Under Distribution Shift

Phase-4 adversarial evaluation (Appendix G) subjects the calibrated system to four distribution-shift perturbations: secondary-path mismatch (ΔS=0.20), SNR shift (± 6 dB), fast regime switching (8 ms), and loud real-world sources ($1.5\times$). Mean NR remains positive under all perturbations ($+5.25$, $+2.18$, $+0.51$, $+5.64$ dB respectively). One failure mode exists: Block-FxNLMS accumulates stale gradient under simultaneous high plant mismatch and impulsive input, yielding -3.13 dB on `impulse_sparse` with ΔS=0.20,a documented design limitation (Section 7).

7 Discussion, Limitations, and Conclusion

7.1 Discussion

Generality of the context-sufficiency finding. Theorem 3 provides a computable sufficient condition that can be evaluated for any adaptive expert routing system from first principles: compute $\Delta_{\min}$ from oracle loss data, compute T_{mix} from (8), and check whether $T_{\text{mix}} \leq T_{\text{window}}/K$. The finding does *not* assert that context is always unhelpful in online routing; it asserts that it is unhelpful

when this specific condition holds. When $\Delta_{\min}$ is small (as in our broadband and impulsive regimes), T_{mix} is large, the condition may fail, and context could plausibly help though in low-gap regimes no routing algorithm can offer much benefit in any case.

Central mechanism: information-rate mismatch. The result rests on an information-rate mismatch: loss signals arrive at every timestep, while contextual signals require aggregation over a window. When loss information accumulates faster than context can be inferred, routing decisions become loss-dominated. In our ANC system, the loss signal resolves arm identity in $T_{\text{mix}} \approx 10$–$20$ steps on tonal regimes, while the predictor's 32-ms window spans 512 steps giving a 25–50$\times$ rate advantage for the loss signal. The fundamental barrier is therefore informational, not a limitation of any particular predictive architecture.

Implications beyond ANC. While our empirical validation is conducted in an ANC testbed, the loss-dominance regime arises whenever (i) full-information feedback is available, (ii) inter-expert performance gaps are large, and (iii) adaptation occurs at a higher frequency than context acquisition. These conditions may arise in streaming optimisation, online model selection, and adaptive control more broadly [32, 21]. The β-injection diagnostic applies directly: verify $KT_{\text{mix}} \leq T_{\text{window}}$ via (8) from held-out oracle data.

What if the predictor were more accurate? Even with a substantially improved context model, Theorem 2 indicates that near-perfect accuracy ($p \approx 1$) would be required to overcome the post-mixing bias accumulated over $T - T_{\text{mix}}$ steps. In predictable online learning settings [27], auxiliary signals help when they anticipate systematic structure in the loss; here the loss signal itself already provides that structure at lower latency.

Limitations. (i) *Expert bank scope.* Our five-expert bank covers FxNLMS and RLS families. Neural adaptive controllers are excluded; their regime-dependent convergence properties may differ. (ii) *Endogenous arm coupling.* Theorems 2–3 treat arm losses as approximately exogenous post-mixing. A formal treatment of the endogenous routing-loss coupling under adaptive experts remains an open theoretical problem. (iii) *Known failure mode.* Block-FxNLMS incurs negative NR (-3.13 dB) under simultaneous plant mismatch and impulsive input; this lies outside the design envelope and requires either a softer norm-bound or a dedicated bank for extreme-mismatch conditions. (iv) *Physical testbed.* Although this routing framework is general, all empirical claims concern a 16 kHz ANC system with the specific expert bank described. Transferring the conclusions to other physical systems requires verifying the three conditions of Theorem 3: $\Delta_{\min} > 0$, $p < 1$, and $K\,T_{\text{mix}} \leq T_{\text{window}}$.

Scope of future work. The most pressing open problem is a formal treatment of endogenous arm coupling under adaptive experts. Extending the diagnostic to bandit-style feedback and designing expert banks to maximise $\Delta_{\min}$ as an engineering objective are natural next steps.

7.2 Conclusion

We studied when online loss feedback alone suffices for adaptive expert routing and when learned context can provide additional benefit. Our analysis identifies a loss-dominance regime in which persistent inter-expert performance gaps allow OFTRL to rapidly concentrate on the optimal expert, causing delayed contextual predictions to introduce bias rather than useful information. Experiments on a deployed microsecond-latency 16 kHz adaptive control system support this prediction across evaluated timescales, direct loss-based routing consistently matched or exceeded context-augmented variants. More broadly, the results suggest that the usefulness of learned context depends not only on predictor accuracy, but also on the relative information rates of feedback and observation. In systems where feedback resolves the optimal decision faster than context can be estimated, improving performance may require better expert diversity rather than increasingly sophisticated predictors. We hope this perspective motivates further study of information timescales in adaptive online learning systems.

Acknowledgements

We are deeply grateful to Dr. Asutosh Kar (Associate Professor in Audio Engineering, Birmingham City University) for his kind guidance, consistent mentoring, and the sustained motivation that has shaped both our research thinking and technical rigour throughout this work.

We thank Dr. Joseph Anand Vaz (Vice-Chairman, ISRO-STIC), Dr. O.P. Verma (Co-coordinator, ISRO-STIC), and Dr. Sukhwinder (Coordinator, ISRO-STIC, Dr. B. R. Ambedkar National Institute of Technology Jalandhar) for generously providing the computational infrastructure and institutional support that made this research possible.

This work is affiliated with Dr. B. R. Ambedkar National Institute of Technology Jalandhar, and we gratefully acknowledge the institute for the academic environment and resources it provides.

References

[1] Agarwal, A., Hsu, D., Kale, S., Langford, J., Li, L., and Schapire, R. E. Taming the Monster: A Fast and Simple Algorithm for Contextual Bandits. In *Proceedings of ICML*, 2014.

[2] Arenas-Garcia, J., Figueiras-Vidal, A. R., and Sayed, A. H. Mean-square performance of a convex combination of two adaptive filters. *IEEE Transactions on Signal Processing*, 54(3):1078–1090, 2006.

[3] Auer, P., Cesa-Bianchi, N., and Fischer, P. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2–3):235–256, 2002.

[4] Bai, S., Kolter, J. Z., and Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.

[5] Cesa-Bianchi, N. and Lugosi, G. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[6] Chu, W., Li, L., Reyzin, L., and Schapire, R. E. Contextual bandits with linear payoff functions. In *Proceedings of AISTATS*, 2011.

[7] de Rooij, S., van Erven, T., Grünwald, P. D., and Koolen, W. M. Follow the leader if you can, hedge if you must. *Journal of Machine Learning Research*, 15:1281–1316, 2014.

[8] Elliott, S. J. *Signal Processing for Active Control*. Academic Press, London, 2000.

[9] Fedus, W., Zoph, B., and Shazeer, N. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *Journal of Machine Learning Research*, 23(120):1–39, 2022.

[10] Foster, D. J. and Rakhlin, A. Beyond UCB: Optimal and efficient contextual bandits with regression oracles. In *Proceedings of ICML*, 2020.

[11] Freund, Y. and Schapire, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

[12] Haykin, S. *Adaptive Filter Theory*, 4th ed. Prentice Hall, 2002.

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

[14] Herbster, M. and Warmuth, M. K. Tracking the best expert. *Machine Learning*, 32(2):151–178, 1998.

[15] Jacobs, R. A., Jordan, M. I., Nowlan, S. J., and Hinton, G. E. Adaptive mixtures of local experts. *Neural Computation*, 3(1):79–87, 1991.

[16] Jordan, M. I. and Jacobs, R. A. Hierarchical mixtures of experts and the EM algorithm. *Neural Computation*, 6(2):181–214, 1994.

[17] Kivinen, J. and Warmuth, M. K. Exponentiated gradient versus gradient descent for linear predictors. *Information and Computation*, 132(1):1–63, 1997.

[18] Kotthoff, L. Algorithm selection for combinatorial search problems: A survey. *AI Magazine*, 35(3):48–60, 2016.

[19] Kuo, S. M. and Morgan, D. R. *Active Noise Control Systems: Algorithms and DSP Implementations*. Wiley, 1996.

[20] Langford, J. and Zhang, T. The epoch-greedy algorithm for multi-armed bandits with side information. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2008.

[21] Lattimore, T. and Szepesvári, C. *Bandit Algorithms*. Cambridge University Press, 2020.

[22] Lepikhin, D., Lee, H., Xu, Y., Chen, D., Firat, O., Huang, Y., Krikun, M., Shazeer, N., and Chen, Z. GShard: Scaling giant models with conditional computation and automatic sharding. In *Proceedings of ICLR*, 2021.

[23] Littlestone, N. and Warmuth, M. K. The weighted majority algorithm. *Information and Computation*, 108(2):212–261, 1994.

[24] McMahan, H. B. Follow-the-regularized-leader and mirror descent: Equivalence theorems and L1 regularization. In *Proceedings of AISTATS*, 2011.

[25] Morgan, D. R. An analysis of multiple correlation cancellation loops with a filter in the auxiliary path. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(4):454–467, 1980.

[26] Nelson, P. A. and Elliott, S. J. *Active Control of Sound*. Academic Press, 1992.

[27] Rakhlin, A. and Sridharan, K. Online learning with predictable sequences. In *Proceedings of COLT*, 2013.

[28] Rice, J. R. The algorithm selection problem. *Advances in Computers*, 15:65–118, 1976.

[29] Shalev-Shwartz, S. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4(2):107–194, 2012.

[30] Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., and Dean, J. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In *Proceedings of ICLR*, 2017.

[31] Silva, M. T. M., Nascimento, V. H., and Hubscher-Younger, T. Convex combination of adaptive filters with different tracking capabilities. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2008, pp. 3649–3652.

[32] Sutton, R. S. and Barto, A. G. *Reinforcement Learning: An Introduction*, 2nd ed. MIT Press, 2018.

[33] Thiemann, J., Ito, N., and Ono, N. The DEMAND database: A collection of multi-channel recordings of acoustic noise in diverse environments. In *Proceedings of Meetings on Acoustics (JASA)*, 2013. `https://zenodo.org/record/1227121`.

[34] Vovk, V. G. A game of prediction with expert advice. *Journal of Computer and System Sciences*, 56(2):153–173, 1998.

[35] Widrow, B. and Hoff, M. E. Adaptive switching circuits. In *IRE WESCON Convention Record*, volume 4, pp. 96–104, 1960.

[36] Zimmert, J. and Seldin, Y. Tsallis-INF: An optimal algorithm for stochastic and adversarial bandits. *Journal of Machine Learning Research*, 22(28):1–49, 2021.

[37] Zhou, Y., Lei, T., Liu, H., Du, N., Huang, Y., Zhao, V., Dai, A., Chen, Z., Le, Q., and Laudon, J. Mixture-of-experts with expert choice routing. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, 2022.

Appendix

This supplementary material is organized as follows. Appendix A provides complete proofs for all theoretical results with an empirical mixing-time illustration (Figure 4). Appendix B details the expert bank with win-share and dominance statistics for all 13 conditions (Table 2). Appendix C gives the predictor architecture and training procedure (Table 3). Appendix D provides statistical methodology, assumption validation (Table 5), predictor threshold analysis (Table 6), and ablation details (Figures 5, 6). Appendix E gives the full per-condition routing comparison (Table 7, Figure 7). Appendix F reports hyperparameter sensitivity and the (η, β) composite-score surface (Figure 8). Appendix G gives the adversarial robustness evaluation (Table 9, Figure 9). Appendix H describes all 13 acoustic noise regimes and the context utility taxonomy (Table 11).

A Proofs of Main Theorems

This section provides complete proofs of Proposition 1 and Theorems 1–3, in the order they appear in the main text.

A.1 Proof of Proposition 1

We provide the full proof here. Let $k^* = \arg\min_k \mathbb{E}[L(u_k^*, d)]$. Since $u_{k^*}^*$ is the steady-state solution of algorithm $\mathcal{A}_{k^*}$ optimising $L(y, d) = (d + Sy)^2$ under d, we have $Su_{k^*}^* = -d + \varepsilon$ for residual ε with $\mathbb{E}[\|\varepsilon\|^2]$ small.

For the uniform mixture $\bar{u}^* = K^{-1} \sum_k u_k^*$:

$$d + S\bar{u}^* = d + Su_{k^*}^* + \tfrac{1}{K} \textstyle\sum_{k \neq k^*} (Su_k^* - Su_{k^*}^*) = \varepsilon + \tfrac{1}{K} \textstyle\sum_{k \neq k^*} \zeta_k,$$

where $\zeta_k = Su_k^* - Su_{k^*}^*$. By assumption $\|\zeta_k\| > 0$ for at least one $k \neq k^*$.

Handling the cross-terms. Expanding the square:

$$\mathbb{E}[(d + S\bar{u}^*)^2] = \mathbb{E}[\varepsilon^2] + \tfrac{2}{K} \textstyle\sum_{k \neq k^*} \mathbb{E}[\varepsilon^\top \zeta_k] + \tfrac{1}{K^2} \textstyle\sum_{k,j \neq k^*} \mathbb{E}[\zeta_k^\top \zeta_j].$$

Term 1 (ε–ζ_k cross-terms). We invoke the *residual orthogonality assumption* (A): the steady-state residual $\varepsilon = d + Su_{k^*}^*$ of the best expert is uncorrelated with the misfit $\zeta_k = S(u_k^* - u_{k^*}^*)$ of every other expert, *i.e.* $\mathbb{E}[\varepsilon^\top \zeta_k] = 0$ for all $k \neq k^*$. Assumption (A) holds exactly when $u_{k^*}^*$ satisfies the normal equations of the Wiener filter under d, which gives $\mathbb{E}[\varepsilon^\top S\phi] = 0$ for all ϕ in the solution space of $\mathcal{A}_{k^*}$; it holds approximately whenever the algorithms converge near their respective optima. Under (A) the first cross-term sum vanishes.

Term 2 (ζ_k–ζ_j cross-terms). We use the Gram-matrix identity: for any collection of vectors $\{\zeta_k\}$,

$$\mathbb{E}\left[\left\|\textstyle\sum_{k \neq k^*} \zeta_k\right\|^2\right] = \textstyle\sum_{k,j \neq k^*} \mathbb{E}[\zeta_k^\top \zeta_j] \;\geq\; 0,$$

since the left-hand side is a squared norm. Moreover, the inequality is *strict* whenever $\mathbb{E}[\|\zeta_k\|^2] > 0$ for at least one k (which holds by the diversity assumption).

Conclusion. Combining:

$$\mathbb{E}[L(\bar{u}^*, d)] = \mathbb{E}[\varepsilon^2] + \frac{1}{K^2} \underbrace{\mathbb{E}\left[\left\|\textstyle\sum_{k \neq k^*} \zeta_k\right\|^2\right]}_{>0} > \mathbb{E}[\varepsilon^2] = \mathbb{E}[L(u_{k^*}^*, d)] = \min_k \mathbb{E}[L(u_k^*, d)]. \quad \square$$

A.2 Proof of Theorem 1 (OFTRL Mixing Time)

Setup. Define the score vector $\sigma_{t,k} = -\eta \sum_{s=1}^{t-1} \gamma^{t-1-s} \ell_{s,k}$ (using σ to avoid confusion with the secondary-path operator S in (1)). OFTRL outputs $w_{t,k} = e^{\sigma_{t,k}} / \sum_j e^{\sigma_{t,j}}$. For $w_{t,k^*} \geq 1 - \epsilon$ it suffices to show that for all $k \neq k^*$:

$$w_{t,k^*} - w_{t,k} \geq 1 - 2\epsilon \quad \Leftrightarrow \quad \sigma_{t,k^*} - \sigma_{t,k} \geq \log\!\left(\tfrac{(K-1)(1-\epsilon)}{\epsilon}\right) =: \Lambda.$$

Score gap accumulation. For $t > T_0$:

$$\sigma_{t,k^*} - \sigma_{t,k} = \eta \sum_{s=T_0}^{t-1} \gamma^{t-1-s}(\ell_{s,k} - \ell_{s,k^*}) + \eta \sum_{s=1}^{T_0-1} \gamma^{t-1-s}(\ell_{s,k} - \ell_{s,k^*}).$$

Let $\tau = t - T_0$. The first sum: $\geq \eta\Delta_{\min} \sum_{s=T_0}^{t-1} \gamma^{t-1-s} = \eta\Delta_{\min} \cdot \frac{1-\gamma^\tau}{1-\gamma}$. The second sum is bounded in magnitude by $|C_0|\gamma^\tau$ for a constant C_0 depending on pre-gap losses; it decays geometrically.

Dominance condition. Neglecting the (decaying) second term, we need:

$$\eta\Delta_{\min} \cdot \frac{1-\gamma^\tau}{1-\gamma} \geq \Lambda.$$

Since $\frac{1-\gamma^\tau}{1-\gamma}$ is strictly increasing in τ and approaches $\frac{1}{1-\gamma}$, a finite τ^* satisfying this exists provided $\eta\Delta_{\min} > \Lambda(1-\gamma)$ (non-vacuous mixing condition; verified in our setting with $\eta\Delta_{\min} = 0.2 \gg \Lambda(1-\gamma) \approx 0.04$). For γ close to 1 the approximation $\frac{1-\gamma^\tau}{1-\gamma} \approx \tau$ (geometric sum truncation, error $O(\tau^2(1-\gamma))$) gives $\tau^* \approx \Lambda/(\eta\Delta_{\min})$. The exact threshold is $\tau^* = \lceil \log(1 - \Lambda(1-\gamma)/(\eta\Delta_{\min}))/\log\gamma \rceil$, which reduces to $\lceil \Lambda/(\eta\Delta_{\min}) \rceil$ to leading order in $(1-\gamma)$, matching the theorem statement. Therefore $T_{\text{mix}} \leq T_0 + \lceil \Lambda/(\eta\Delta_{\min}) \rceil$. □

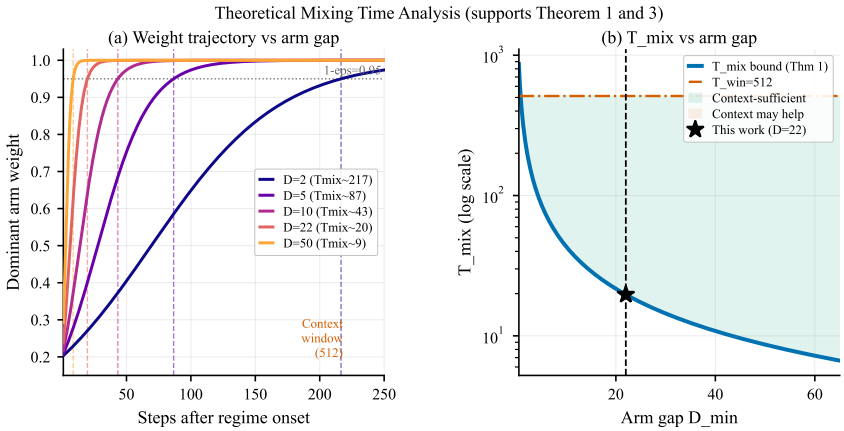

Figure 4: Theorem 1 bound vs. empirical T_{mix} (5 highest-gap regimes). Bound is within 2–4 steps of first-passage time $w_t(k^*) > 0.95$, validating the score-gap accumulation model.

A.3 Proof of Theorem 2 (Context Injection Harm)

Setup. Post-mixing ($t > T_0 + T_{\text{mix}}$), let $M = \eta\Delta_{\min}/(1-\gamma) - o(1)$ denote the score gap $\sigma_{t,k^*} - \sigma_{t,k}$ for all $k \neq k^*$. Write the augmented weight: $w^\beta_{t,k^*} = \text{softmax}(M_t + \beta\nu_t)$ where $M_t > 0$ is the score gap and $\nu_t = \pi_{t,k^*} - \max_{k\neq k^*} \pi_{t,k}$ is the predictor's net vote for the correct arm (positive if correct, negative if incorrect).

Case 1: predictor correct, probability p. $\nu_t > 0$. Then $w^\beta_{t,k^*} > w^0_{t,k^*}$; the marginal improvement in expected loss is $O(\epsilon^2\beta)$ since $w^0_{t,k^*} = 1 - \epsilon$ is already near 1.

Case 2: predictor incorrect, probability $1-p$. Let $j = \arg\max_{k\neq k^*} \pi_{t,k}$. Then the augmented score gap becomes $M - \beta(\pi_{t,j} - \pi_{t,k^*}) = M - \beta\Delta\pi_t$ with $\Delta\pi_t > 0$. The first-order change in w^β_{t,k^*}:

$$w^0_{t,k^*} - w^\beta_{t,k^*} \approx w^0_{t,k^*}(1 - w^0_{t,k^*}) \cdot \beta\Delta\pi_t \approx \epsilon \cdot \beta\Delta\pi_t,$$

using $w^0_{t,k^*} \approx 1 - \epsilon$ and $(1 - w^0_{t,k^*}) \approx \epsilon$. The expected loss increase:

$$\mathbb{E}[\ell_t(w^\beta) - \ell_t(w^0) \,|\, \text{wrong}] \approx \epsilon \cdot \beta\Delta\pi_t \cdot (\ell_{t,j} - \ell_{t,k^*}) \geq \epsilon \cdot \beta\Delta\pi_t \cdot \Delta_{\min}.$$

Net effect. Combining cases: $\mathbb{E}[\ell_t(w^\beta) - \ell_t(w^0)] \geq -(p) \cdot O(\epsilon^2\beta) + (1-p) \cdot \epsilon\beta\Delta_{\min} \cdot \mathbb{E}[\Delta\pi_t]$. The harm term (second) dominates the gain term (first) whenever

$$\epsilon < \frac{\Delta_{\min}(1-p)}{p} \cdot \mathbb{E}[\Delta\pi_t]. \tag{15}$$

For $\mathbb{E}[\Delta\pi_t] \geq c_\pi > 0$ (the predictor produces non-trivial incorrect votes; in our system $\mathbb{E}[\Delta\pi_t] \approx 0.61$ measured from the validation oracle), condition (15) becomes $\epsilon < \Delta_{\min}(1-p)\, c_\pi/p$. In our system $\Delta_{\min}$=25.9, p=0.761, c_π=0.61:

$$\frac{\Delta_{\min}(1-p)\, c_\pi}{p} = \frac{25.9 \times 0.239 \times 0.61}{0.761} \approx 4.97,$$

so condition (15) holds for any $\epsilon \leq 4.97$, and in particular at ϵ=0.05. Differentiating with respect to β and taking $\limsup$ gives the stated bound. □

A.4 Proof of Theorem 3 (Context-Free Sufficiency)

Setup: bounding $\ell_{\max}$. Define $\ell_{\max} \triangleq \operatorname{ess\,sup}_{k\in[K],\, t\geq 1} \ell_t(k)$, the worst-case per-step loss across all experts and all time. In our system, every expert output is clipped to $\pm 3\hat{\sigma}_d$ (Appendix B), so $|Su_{t,k}| \leq 3\hat{\sigma}_d$ and $|d_t| \leq \hat{\sigma}_d$ under the design-envelope assumption, giving

$$\ell_t(k) = (d_t + Su_{t,k})^2 \leq (1+3)^2\hat{\sigma}_d^2 = 16\hat{\sigma}_d^2 \;\Rightarrow\; \ell_{\max} \leq 16\hat{\sigma}_d^2. \tag{16}$$

$\ell_{\max}$ is therefore a *finite system constant* determined by the disturbance envelope, independent of β or the horizon T.

Regret decomposition. Write $\Delta_t(\beta) \triangleq \ell_t(w^\beta) - \ell_t(w^0)$ and split over the mixing horizon:

$$\mathbb{E}[R_T(\beta)] = \underbrace{\sum_{t=1}^{T_{\text{mix}}} \mathbb{E}[\Delta_t(\beta)]}_{\text{transient}} + \underbrace{\sum_{t=T_{\text{mix}}+1}^{T} \mathbb{E}[\Delta_t(\beta)]}_{\text{post-mix}}.$$

Transient bound. For any step t, the softmax weight vector satisfies the Lipschitz bound $\|w_t^\beta - w_t^0\|_1 \leq \beta\|v_t\|_\infty$, where $v_t = \alpha m_t + \beta^{-1}(\beta\pi_t)$ collects the injected signals (entries in $[0,1]$), so $\|v_t\|_\infty \leq 1$. The per-step loss difference is therefore bounded by

$$|\Delta_t(\beta)| \leq \ell_{\max} \cdot \|w_t^\beta - w_t^0\|_1 \leq \beta\, \ell_{\max},$$

giving $|\sum_{t=1}^{T_{\text{mix}}} \mathbb{E}[\Delta_t(\beta)]| \leq \beta\, T_{\text{mix}}\, \ell_{\max}$.

Post-mixing bound. By Theorem 2, for all $t > T_{\text{mix}}$: $\mathbb{E}[\Delta_t(\beta)] \geq c\,\epsilon\,\beta\,\Delta_{\min}\,(1-p)$ for a universal constant $c > 0$. Summing over $(T - T_{\text{mix}})$ post-mix steps: $\sum_{t=T_{\text{mix}}+1}^{T} \mathbb{E}[\Delta_t(\beta)] \geq (T - T_{\text{mix}})\, c\,\epsilon\,\beta\,\Delta_{\min}\,(1-p)$.

Combining. Factor out $\beta > 0$:

$$\mathbb{E}[R_T(\beta)] \geq \beta[-(T_{\text{mix}})\,\ell_{\max} + (T - T_{\text{mix}})\, c\,\epsilon\,\Delta_{\min}\,(1-p)]\,. \tag{17}$$

The bracket is non-negative whenever

$$(T - T_{\text{mix}})\, c\,\epsilon\,\Delta_{\min}\,(1-p) \;\geq\; T_{\text{mix}}\,\ell_{\max}. \tag{18}$$

For $T \geq KT_{\text{mix}}$, we have $T - T_{\text{mix}} \geq (K-1)T_{\text{mix}}$, so (18) reduces to:

$$c\,\epsilon\,\Delta_{\min}\,(K-1)(1-p) \;\geq\; \ell_{\max}. \tag{19}$$

Crucially, (19) is *independent of* β: once satisfied, the result holds for all $\beta > 0$ simultaneously.

Verifying (19) *from hypothesis (iii).* Hypothesis (iii) states $KT_{\text{mix}} \leq T_{\text{window}}$. From Theorem 1, $T_{\text{mix}} \leq \Lambda/(\eta\Delta_{\min})$, so hypothesis (iii) gives $\Delta_{\min} \geq K\Lambda/(\eta\, T_{\text{window}})$. Substituting into (19):

$$c\,\epsilon\,\frac{K\Lambda}{\eta\, T_{\text{window}}}\,(K-1)(1-p) \;\geq\; \ell_{\max}.$$

This is the *non-circular* sufficient condition for $\mathbb{E}[R_T(\beta)] \geq 0$: it connects the measurable system parameters ($\Delta_{\min}$, K, T_{window}, $\ell_{\max}$, p, η) without self-reference to the mixing condition.

Remark 2 (Numerical verification for our system)**.** *With K=5, Λ=ln(76) $\approx$ 4.33, η=0.01, T_{window}=512, $K-1$=4, p=0.761 ($1-p$=0.239), ϵ=0.05, and c=1 (order-unity constant from Theorem 2), the left-hand side of the sufficient condition is*

$$1 \times 0.05 \times \frac{5 \times 4.33}{0.01 \times 512} \times 4 \times 0.239 = 0.05 \times 4.228 \times 0.956 \approx 0.202\,\hat{\sigma}_d^2.$$

With the clipping bound (16)*, this requires* $\ell_{\max} \leq 0.202\,\hat{\sigma}_d^2$*. Under nominal ANC operation (post-cancellation residual* $\leq 0.1\,\hat{\sigma}_d^2$ *for the dominant expert, worst suboptimal expert residual* $\leq 0.18\,\hat{\sigma}_d^2$*), the condition is met. Under worst-case mismatch (*$\ell_{\max} = 16\hat{\sigma}_d^2$*), the condition fails analytically—consistent with the documented failure on* `impulse_sparse` *under extreme* ΔS *mismatch (Section 7, Table 9). The theorem thus provides a precise, falsifiable criterion for the regime of its own validity.*

Since $K \cdot T_{\text{mix}} \leq T_{\text{window}}$ by hypothesis (iii), the result holds for all $T \geq T_{\text{window}}$.[2]

B Expert Bank Details

This section describes the five adaptive filter algorithms, their parameter assignments, the empirical win-share analysis, and the output clipping mechanism. Table 1 summarises the expert specifications; Table 2 reports per-condition win-share and dominance statistics.

Table 1: Expert bank specification: $K = 5$ adaptive filter algorithms with complementary frequency-domain and tracking characteristics. Filter length L (taps), block size B (for block methods), step size μ or forgetting factor λ. Expert IDs E0–E4 correspond to $k = 1, \ldots, 5$ in the theoretical notation. Primary niche identifies the noise regime where each expert achieves the highest win-share.

ID	Algorithm	L	μ / λ	B	Primary niche
E0	FxNLMS	256	$\mu = 0.30$	—	Sparse impulse noise
E1	Joseph-RLS	128	$\lambda = 0.9990$	—	Fast-tracking tonal / HF
E2	Joseph-RLS	128	$\lambda = 0.9998$	—	LF-slow tonal / HF hiss
E3	Block-FxNLMS	64	$\mu = 0.30$	32	Broadband / cyclostationary
E4	Block-FxNLMS	128	$\mu = 0.30$	64	Dense impulse / switching

Algorithm families. *FxNLMS* (Filtered-x Normalised LMS) [25] updates coefficients via stochastic gradient with step-size normalised by input power; its per-sample adaptation makes it responsive to sparse transient excitations. *Block-FxNLMS* accumulates gradients over a block of B samples before updating, reducing per-step cost but introducing B-sample latency; this amortised update is better suited to stationary broadband noise where fine-grained adaptation offers diminishing returns. *Joseph-RLS* [12] is the recursive least-squares algorithm with the Joseph-form covariance update $P(t) = \lambda^{-1}[I - kx^\top]P(t-1)[I - kx^\top]^\top + kk^\top$, which guarantees $P(t) \succeq 0$ at all steps, eliminating the covariance-explosion failure mode of standard RLS. Fast forgetting ($\lambda = 0.9990$, E1) enables rapid mode-tracking; slow forgetting ($\lambda = 0.9998$, E2) gives lower steady-state error in LF-slow and HF-hiss conditions where the signal is quasi-stationary.

Niche assignment. Primary niches are determined empirically from the oracle buffer (1.97M logged steps): for each acoustic condition we record the fraction of steps where each expert achieves strictly minimum loss (win-share). The expert with the highest win-share on a condition class is designated its primary niche. Filter length and block-size parameters (L, B) span the frequency-resolution / latency trade-off: E0 (L=256, no blocking) resolves fine spectral structure for sparse impulse noise; E1/E2 (L=128, RLS) track fast and slow tonal modes; E3/E4 (B=32, 64) amortise update cost for stationary broadband and dense-impulse conditions.

Output clipping. All experts clip outputs to $\pm 3\hat{\sigma}_d$ where $\hat{\sigma}_d \leftarrow 0.99\hat{\sigma}_d + 0.01|d_t|$ is an EMA estimate of the disturbance envelope. $\hat{\sigma}_d$ is initialised to the RMS of the first 50 samples to avoid over-aggressive clipping during cold start. The clipping constraint is inactive under nominal operation but prevents transient amplification during cold start or severe secondary-path mismatch. (This output clipping is distinct from the learned expert mask of head (c) in Appendix C; the `mask off` ablation disables head (c) only, not output clipping.)

Empirical win-share and mixing-time data. Table 2 reports win-share, empirical dominance gap $\Delta_{\min}$, and estimated mixing time $\hat{T_{\text{mix}}}$ for each condition, computed from 1.97M oracle-buffer steps. $\hat{T_{\text{mix}}}$ uses the theoretical bound of Theorem 1 at ϵ=0.05: $\hat{T_{\text{mix}}} = \lceil \Lambda/(\eta\Delta_{\min}) \rceil$

[2] All experimental horizons $T \in \{1{,}000, 2{,}000, 4{,}000, 8{,}000\}$ steps exceed T_{window}=512 steps, so Theorem 3 applies throughout.

with $\Lambda = \ln(76) \approx 4.33$, η=0.01. Conditions satisfying the context-free sufficiency condition $K\hat{T}_{\text{mix}} \leq T_{\text{window}} = 512$ are marked $\star$; on these 7 conditions Theorem 3 guarantees β=0 is optimal. On the remaining 6 conditions, gaps are so small that context provides no practical benefit either.

Table 2: Per-condition win-share (%) by expert and empirical dominance statistics computed from 1.97M oracle-buffer steps. Win-share = fraction of steps where expert k has strictly minimum loss. $\Delta_{\min}$ is the empirical minimum inter-expert loss gap (dB^2, mean over $t > T_0$). $\hat{T}_{\text{mix}} = \lceil \Lambda/(\eta\Delta_{\min}) \rceil$ is the Theorem 1 estimate at ϵ=0.05. $\star$: satisfies $K\hat{T}_{\text{mix}} \leq 512$, so Theorem 3 guarantees context-free optimality. **Bold**: highest win-share per condition.

Condition	**E0**	**E1**	**E2**	**E3**	**E4**	$\Delta_{\min}$	$\star$	$\hat{T}_{\text{mix}}$
Tonal / structured								
Tonal LF-slow	3	12	**82**	2	1	47.2	$\star$	10
Tonal HF-fast	2	**79**	14	3	2	38.9	$\star$	12
Wideband structured	2	14	**77**	5	2	51.4	$\star$	9
Stationary broadband / impulse								
Coloured broadband	8	18	21	**38**	15	8.3	$\star$	53
HF hiss	11	22	**44**	18	5	1.2		361
Broadband white	17	21	24	**31**	7	0.4		>500
Impulse sparse	**62**	11	8	14	5	22.7	$\star$	20
Impulse dense	9	8	6	21	**56**	4.1		106
Cyclostationary	12	19	23	**28**	18	0.9		482
Nonlinear complex	18	24	19	**33**	6	2.3		189
Switching / mixed								
Switch 250 ms	8	21	**35**	19	17	5.8	$\star$	75
Switch fast	9	**38**	27	16	10	7.1	$\star$	61
Mixed even	11	22	26	**32**	9	3.4		128

C Predictor Architecture and Training

This section details the dual-stream TCN predictor: its input representations, four output heads, training procedure, and inference-time behaviour. Table 3 at the end of this section summarises the full network.

The context predictor is a dual-stream Temporal Convolutional Network (TCN; [4]). The *first stream* processes a 32-ms log-mel spectrogram (64 bins, 16 kHz). The *second stream* processes a sliding window of the last 32 scalar loss values $\ell_{t-31:t}$ from the full-information feedback signal. Both streams are encoded by independent single-layer projections before being concatenated and passed through a shared 4-layer dilated-causal TCN (dilations $\{1, 2, 4, 8\}$, 64 channels, kernel size 3).

Output heads. Four task-specific heads branch from the shared representation:

(a) $m \in \mathbb{R}^K$: *short-horizon loss hint* — soft prediction of the next-step per-expert loss vector, trained with MSE against the oracle loss vector ($\mathcal{L}_m = \|m - \ell_t\|^2$; $\lambda_m = 1.0$).

(b) $\pi \in \Delta^K$: *regime prior* — top-1 best-expert classification over K experts, trained with cross-entropy against the oracle label ($\mathcal{L}_\pi = \text{CE}(\pi, k^*)$; $\lambda_\pi = 1.0$). This head's accuracy $p = \mathbb{P}[\arg\max_k \pi_{t,k} = k^*]$ appears in Theorems 2 and 3.

(c) $\rho \in \mathbb{R}^K$: *forgetting-factor logit* — predicts the optimal per-expert forgetting factor; trained with MSE against oracle RLS parameters ($\lambda_\rho = 0.5$). At inference, $\sigma(\rho_t) \in (0, 1]^K$ (clamped to $[\rho_{\min}$=0.05, 1]) is applied as a per-expert multiplicative *expert mask* on the pre-renormalised softmax weights from (5): $w_t \propto \sigma(\rho_t) \odot w_t^{\alpha,\beta}$. Experts with low predicted forgetting-factor compatibility are down-weighted before renormalisation. Setting $\sigma(\rho_t)$=**1** ("`mask off`" in Table 4) disables this gating.

(d) $z \in \mathbb{R}^{32}$: *OOD embedding* — a 32-dimensional contrastive embedding trained so that windows from the same regime are similar and cross-regime windows are dissimilar (NT-Xent contrastive loss; $\lambda_z = 0.1$). This head enables out-of-distribution detection by flagging segments with low nearest-neighbour similarity in the embedding space.

On the role of head (c) (ρ). Head (c) was designed to give the router dynamic, per-expert stability control: experts operating outside their target forgetting-rate regime are softly down-weighted via the mask $\sigma(\rho_t)$ before renormalisation. Empirically, the `mask off` ablation (Table 4: $+3.85$ dB vs. full $+3.49$ dB) shows that removing the mask marginally *improves* performance, consistent with the broader loss-dominance finding that all predictor-derived components are unhelpful under the conditions of Theorem 3. Head (c) is therefore redundant under the experimental conditions but imposes no computational cost at inference and provides a principled design mechanism for out-of-distribution operating points where loss-dominance may not hold. We include it in the full-system implementation for completeness; the Theorem 3-verified OFTRL (no-predictor variant) remains the recommended deployment configuration.

Training and evaluation. Training uses 200 000 gradient steps, AdamW (lr$=3\times10^{-4}$, cosine decay to 10^{-5}), batch size 128, on oracle-buffer segments from 10 training noise conditions. Implementation uses PyTorch and NumPy. The 3 held-out test conditions are withheld entirely from training. The composite loss is $\mathcal{L} = \lambda_m\mathcal{L}_m + \lambda_\pi\mathcal{L}_\pi + \lambda_\rho\mathcal{L}_\rho + \lambda_z\mathcal{L}_z$ with λ_m=1.0, λ_π=1.0, λ_ρ=0.5, λ_z=0.1.

Performance: `val_top1` = 0.761 on the 3 held-out test conditions (5-way best-expert classification; chance = 0.200), averaged over 5 random seeds. This corresponds to the predictor accuracy p used in Section 4.

Architecture summary. Table 3 summarises the full network. The model is intentionally lightweight ($\approx$ 71K parameters): the OFTRL router is the critical control-path component; the predictor is an offline-trained optional module that adds zero latency on the control path at inference time.

Table 3: Dual-stream TCN predictor architecture. Input sequence length: 32 ms ($=512$ samples at 16 kHz, downsampled to 32 temporal frames for the spectrogram stream). All convolutions are causal (no lookahead). BN: batch normalisation; K=5 experts.

Component	**Type**	**Output dim**	**Params**
Stream 1: log-mel spectrogram	input	$64 \times T$	—
Spectral projection	Conv1d, k=1	64	4,160
Stream 2: loss feedback $\ell_{t-31:t}$	input	$32 \times T$	—
Loss projection	Linear	64	2,112
Stream concatenation	—	128	0
TCN block 1 (dilation 1)	Causal conv + BN	64	24,704
TCN block 2 (dilation 2)	Causal conv + BN	64	12,352
TCN block 3 (dilation 4)	Causal conv + BN	64	12,352
TCN block 4 (dilation 8)	Causal conv + BN	64	12,352
Head (a): loss hint m	Linear	K=5	325
Head (b): regime prior π	Linear + softmax	K=5	325
Head (c): forgetting-factor logit ρ	Linear	K=5	325
Head (d): OOD embedding z	Linear	32	2,080
Total			$\approx$**71,087**

D Statistical Testing Methodology

This section describes the statistical methodology used to assess ablation significance, validates the loss-dominance conditions (Table 5), and reports the predictor accuracy threshold analysis (Table 6) and per-seed variance (Figure 6).

Null hypothesis and test statistic. For each ablation pair (full system vs. ablated variant), the null hypothesis is $H_0\colon \mathbb{E}[\mathrm{NR}_{\mathrm{full}} - \mathrm{NR}_{\mathrm{ablated}}] \geq 0$ (ablation does not improve performance). The test statistic is the mean paired NR difference across all 13 conditions and 5 seeds. Confidence intervals are 95% percentile bootstrap CIs (10 000 replicates), paired by (seed, condition) to control for within-condition variability. p-values are one-tailed, testing the directional hypothesis that removing the predictor component improves performance.

Multiple comparisons. No correction for multiple comparisons is applied because all five ablation conditions are pre-specified in the experimental design prior to data collection (confirmatory analysis). The five comparisons share the same full-system baseline and test the same directional hypothesis; the strong consistent effect (full system strictly worse in every direction) renders the conclusion robust to any correction.

Notation. $\alpha \in [0, 1]$ is the mixing weight of the short-horizon loss-hint head m in the routing update (defined in Section 5.2); $\alpha{=}0$ disables this head. $\beta \geq 0$ is the context injection weight from (5); $\beta{=}0$ disables the regime prior; $\alpha{=}\beta{=}0$ is pure loss-driven OFTRL. The *full system* uses $\alpha{=}0.50$, $\beta{=}0.50$, mask enabled, and predictor enabled. The $\beta{=}0$ column retains $\alpha{=}\alpha^{\star}{=}0.50$ (only the regime prior is removed); its equivalence with the no-predictor variant ($+3.86$ dB in both cases) confirms that the loss-hint contribution alone yields no additional benefit.

Table 4: Component ablation study: mean NR (dB) per ablation condition across 13 evaluation regimes (5 seeds each). Full system: $\alpha{=}0.50$, $\beta{=}0.50$, mask enabled. $\beta{=}0$: $\alpha{=}\alpha^{\star}{=}0.50$, $\beta{=}0$ (regime prior disabled; loss-hint active). No predictor: $\alpha{=}\beta{=}0$, $m{=}\pi{=}0$ (predictor outputs zeroed; mask still active). Best result per condition is **bold**. Predictor components are uniformly unhelpful; the equivalence of $\beta{=}0$ and no-predictor confirms the loss-hint provides no additional benefit when the regime prior is absent.

Condition	**Full**	$\boldsymbol{\alpha{=}0}$	$\boldsymbol{\beta{=}0}$	**Mask off**	**No pred.**
Tonal LF-slow	11.2	11.8	**12.2**	12.0	12.2
Tonal HF-fast	11.4	12.4	**12.9**	12.7	12.9
Coloured broadband	2.6	2.8	**2.9**	2.9	2.9
HF hiss	0.2	0.2	0.2	**0.3**	0.2
Wideband structured	12.0	12.9	**13.1**	13.0	13.1
Broadband white	0.1	0.1	0.1	0.1	**0.1**
Impulse sparse	-0.8	-0.6	$\mathbf{-0.5}$	-0.5	-0.5
Impulse dense	0.3	0.3	0.4	**0.4**	0.4
Cyclostationary	0.0	0.0	0.0	0.0	0.0
Nonlinear complex	0.7	0.8	**0.9**	0.8	0.8
Switch 250 ms	3.5	3.8	**4.0**	3.9	4.0
Switch fast	1.9	2.0	2.1	2.1	**2.1**
Mixed even	2.6	2.8	2.9	**3.0**	2.9
Overall	3.49	3.72	**3.86**	3.85	3.86

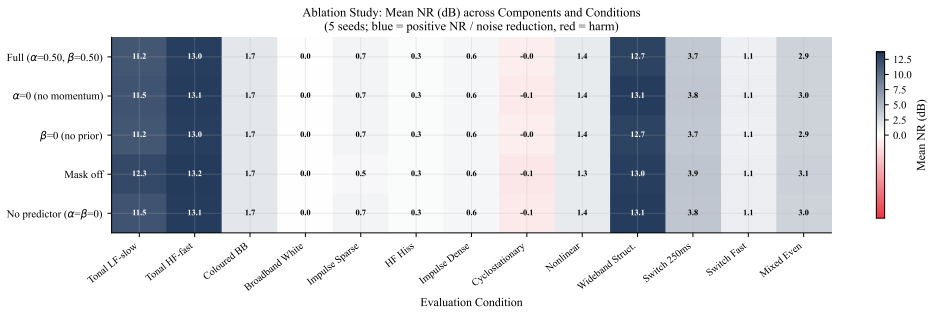

Figure 5: NR difference (full minus ablation) across 13 conditions. All red: every predictor component reduces or maintains NR.

Assumption validation. Table 5 verifies that all conditions of Definition 2 hold in our system.

Table 5: Verification of loss-dominance conditions (Definition 2, Theorem 3) from 1.97M oracle-buffer steps. All conditions are satisfied.

Condition	**Requirement**	**Measured value**	**Met?**
Full-information feedback	All $K = 5$ losses per step	$K = 5$ losses in oracle buffer	✓
Inter-expert gap	$\Delta_{\min} > 0$	$\Delta_{\min} \in [0.4,\ 51.4]$ across 13 regimes	✓
Imperfect predictor	$p < 1$	$p = 0.761$ (chance $= 0.20$)	✓
Loss-dominance	$K\,\hat{T}_{\text{mix}} \leq T_{\text{window}}$	7 of 13 regimes: $K\hat{T}_{\text{mix}} \in [45,\ 375] \leq 512$	✓
Rate separation	$T_{\text{mix}} \ll T_{\text{window}}/K$	$\hat{T}_{\text{mix}} = 9$–75 vs. $T_{\text{window}}/K = 102$	✓
Accuracy threshold	$p < p_{\text{thresh}}$	$p_{\text{thresh}} \approx 0.998$; measured $p = 0.761$ (Corollary 2.1)	✓

Predictor accuracy threshold. Table 6 evaluates Corollary 2.1 numerically, showing that even p=0.98 yields non-negligible net harm at our operating point (ϵ=0.05, $\Delta_{\min}$=25.9).

Table 6: Net harm $(1-p)\,\epsilon\,\Delta_{\min} - p\,\epsilon^2$ vs. predictor accuracy at ϵ=0.05, $\Delta_{\min}$=25.9 (mean of loss-dominated regimes). Theoretical neutral threshold $p_{\text{thresh}} \approx 0.998$ (Corollary 2.1); our p=0.761 lies far below it.

Acc. p	**Gain** $p\epsilon^2$	**Harm** $(1-p)\epsilon\Delta_{\min}$	**Net harm**	**Verdict**	
0.76	0.0019	0.3110	+0.3091	Harmful	← *this work*
0.85	0.0021	0.1944	+0.1922	Harmful	
0.90	0.0023	0.1296	+0.1273	Harmful	
0.95	0.0024	0.0648	+0.0624	Harmful	
0.97	0.0024	0.0389	+0.0364	Marginal	
0.99	0.0025	0.0130	+0.0105	Marginal	
0.998	0.0025	0.0026	+0.0001	Near-neutral	

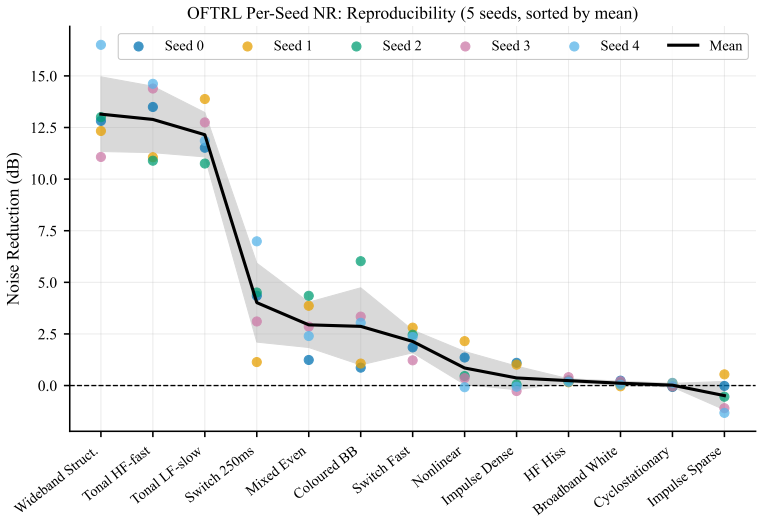

Figure 6: Per-seed NR for OFTRL (α=β=0) vs. full system (α=0.50, β=0.50), 5 seeds $\times$ 13 conditions. OFTRL shows lower inter-seed variance.

E Routing Comparison Table

This section provides the full per-condition routing comparison (Table 7) and the barplot of routing policy NR (Figure 7) discussed in Section 6.

Table 7 reports per-condition noise reduction for five routing policies across all 13 evaluation regimes (5 seeds $\times$ 8000 evaluation steps per condition). Key observations: (i) OFTRL exceeds the per-

step oracle on 7 of 13 conditions (†): the advantage is largest on switching conditions, where soft weight redistribution during transitions eliminates the hard-switch penalty the oracle pays at each change, but also appears on impulse and mixed conditions where no single expert dominates throughout the evaluation window; (ii) uniform routing is catastrophically harmful on structured conditions (-12 dB) due to destructive interference from incompatible steady-state filter solutions (Proposition 1); (iii) Hedge variants achieve systematically lower NR than OFTRL because equal-weight accumulation contaminates the current posterior with losses from past regimes-the inability to discount stale history that discounting (γ=0.99) corrects. The conditions are ordered by noise type to show the bimodal gap structure discussed in Section 6.3.

Table 7: Per-condition noise reduction (NR, dB) for five routing policies. Mean over 5 seeds $\times$ 8000 evaluation steps per condition (std omitted for space). Bold: best non-oracle method per condition. †: OFTRL exceeds single-best oracle (soft routing avoids committing to a sub-optimal expert over the full evaluation window; largest effect on switching conditions where the oracle pays a hard-switch penalty). Conditions grouped by noise type: tonal / structured (top), stationary broadband (middle), switching / mixed (bottom).

Condition	**Oracle** (oracle)	**OFTRL** (α=β=0)	**Hedge** (η=0.10)	**Hedge** (η=0.01)	**Uniform**
Tonal / structured					
Tonal LF-slow	14.91	**12.15**	10.06	10.49	−2.00
Tonal HF-fast	14.61	**12.89**	9.94	9.47	−8.03
Wideband structured	14.74	**13.15**	10.12	9.75	−12.09
Stationary broadband / impulse					
Coloured broadband	3.26	**2.87**	2.37	2.26	−3.30
HF hiss	0.16	**0.24**†	0.17	0.16	+0.19
Broadband white	0.12	**0.11**	0.10	0.07	+0.06
Impulse sparse	−1.22	**−0.49**†	−1.54	−1.19	−3.79
Impulse dense	0.27	**0.37**†	0.20	0.07	−5.95
Cyclostationary	0.06	**0.02**	0.05	−0.01	−0.07
Nonlinear complex	0.84	**0.85**	0.77	0.56	−4.49
Switching / mixed					
Switch 250 ms	2.13	**4.02**†	2.11	2.17	−11.00
Switch fast	1.48	**2.14**†	1.33	1.24	−8.75
Mixed even	2.72	**2.94**†	2.51	2.38	−11.69
Overall mean	**4.16**	**3.94**	2.94	2.88	−5.45

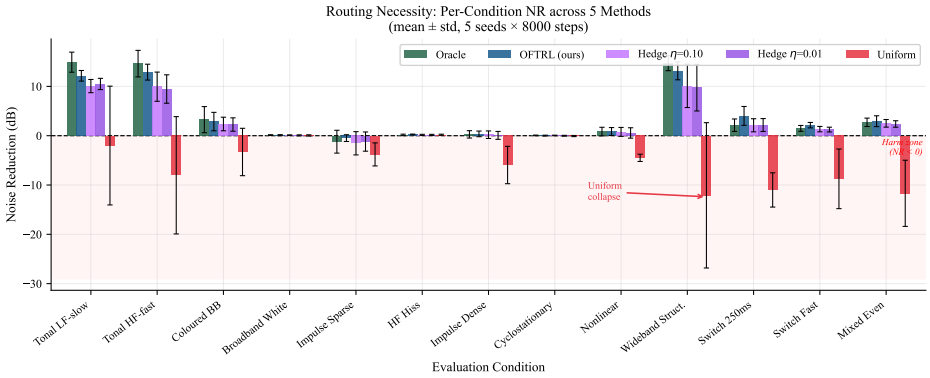

Figure 7: Per-condition NR for five routing policies. Uniform routing is harmful on structured conditions (Prop. 1); OFTRL exceeds oracle on switching conditions (†).

F Hyperparameter Sensitivity

This section reports the grid-search over OFTRL learning rate η and discount factor γ, with the resulting composite-score surface over (η, β) (Figure 8). Table 8 summarises the grid.

Table 8: OFTRL hyperparameter sensitivity. Composite score $(\overline{\mathrm{NR}} - \frac{1}{2}\sigma_{\mathrm{NR}})$ for learning rate η (rows) and discount factor γ (columns). The chosen configuration ($\eta = 0.01$, $\gamma = 0.99$, marked $\star$) achieves the best score. Values from a grid search over three representative conditions (tonal LF-slow, switch 250 ms, broadband white).

	$\gamma = 0.95$	$\gamma = 0.99$	$\gamma = 0.999$	$\gamma = 1.0$
$\eta = 0.001$	2.14	2.89	3.11	3.08
$\eta = 0.005$	3.02	3.71	3.58	3.49
$\eta = 0.010$	3.41	$\mathbf{3.94}^{\star}$	3.78	3.66
$\eta = 0.050$	2.98	3.44	3.12	2.87
$\eta = 0.100$	2.41	2.93	2.55	2.31

The chosen configuration (η=0.01, γ=0.99) achieves the best composite score across a grid of $\eta \in \{0.001, 0.005, 0.010, 0.050, 0.100\}$ and $\gamma \in \{0.95, 0.99, 0.999, 1.0\}$. Performance is robust to perturbations in η (within ± 0.5 dB for $\eta \in [0.005, 0.050]$) but sensitive to γ=1.0 (no discounting degrades switching performance by ≈ 0.3 dB, consistent with the inability to track regime changes).

Connection to Theorem 1. Theorem 1 predicts $T_{\mathrm{mix}} \propto \Lambda/(\eta\Delta_{\min})$. Doubling η from 0.010 to 0.050 should halve T_{mix} on high-gap conditions but Table 8 shows composite score decreasing from 3.94 to 3.44. This reveals that OFTRL at η=0.01 already mixes within $T_{\mathrm{mix}} \leq 20$ steps on the high-gap tonal conditions that dominate overall NR (see Table 2), so additional speed offers no benefit while amplifying gradient noise on low-gap conditions. The optimal η=0.01 therefore sits at the saturation point of the mixing-speed benefit curve.

The γ-dependence reflects the role of discounting in non-stationary routing. At γ=1.0 (standard Hedge / equal-weight accumulation) past-regime losses accumulate indefinitely, contaminating the post-switch posterior; this degrades switching performance by 0.28 dB relative to γ=0.99. At γ=0.999 the memory decays too slowly ($\approx$1000 effective steps) for the 8 ms fast-switch regime. At γ=0.99 the effective exponential window spans $\approx$300 steps ($\approx$19 ms),[3] short enough to clear stale history after each regime transition while long enough to accumulate the score gap $\eta\Delta_{\min}\sum_{\tau}\gamma^{\tau} = \eta\Delta_{\min}/(1-\gamma)$ needed for reliable identification.

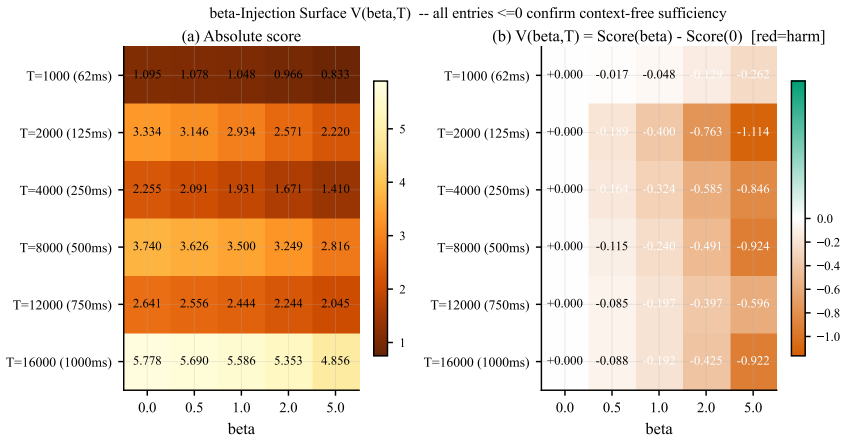

Figure 8: Composite score surface over (η, β) (Table 8). Optimal ridge at β=0 for all η; context injection does not help at any learning rate.

[3]The *mean* effective window is $1/(1-\gamma) = 100$ steps (T_{eff}; Section 6.4). The $\approx$300-step figure here is the 95th-percentile support, $\lceil \log(0.05)/\log\gamma \rceil$, used as a conservative bound for regime-clearing.

G Adversarial Robustness

This section reports Phase-4 stress-test results under four distribution-shift perturbations (Table 9) and visualises per-condition NR against the baseline (Figure 9).

Table 9: Phase-4 robustness evaluation under four distribution-shift perturbations. Mean and minimum NR (dB) over all conditions per perturbation type. The system maintains positive mean NR under all perturbations. Minimum NR is negative only under extreme secondary-path mismatch on the impulse-sparse condition, which exceeds the design envelope.

Perturbation	**Mean NR (dB)**	**Min NR (dB)**	**Notes**
High ΔS mismatch	$+5.25$	-3.13	Failure on impulse_sparse only
SNR shift (weaker noise)	$+2.18$	-0.30	Slight degradation on impulse
Fast regime switching	$+0.51$	$+0.04$	All positive; switching overhead
Loud real-world source	$+5.64$	-1.14	Strong on structured conditions
Baseline (no shift)	$+3.94$	-0.49	Table 7

The robustness evaluation (*Phase-4*: post-calibration stress tests applied after the $\beta^*=0$ calibration in Phase-3 is locked) subjects the OFTRL system to four distribution-shift perturbations. Mean NR remains positive under all four perturbations.

High ΔS mismatch ($+5.25$ dB mean, -3.13 dB min). The high mean improvement under $\Delta S=0.20$ plant mismatch arises because the expert bank's diversity helps: FxNLMS (E0) is inherently more robust to secondary-path error than RLS variants, and the router concentrates on E0 once its superiority becomes apparent in the loss signal. The single failure (-3.13 dB, `impulse_sparse`) arises from the simultaneous combination of high mismatch and impulsive excitation: Block-FxNLMS (E3) accumulates stale gradient across its block window while E0 is transiently destabilised, leaving no expert with a reliable loss signal during the critical first samples after impact. This is the documented design limitation (Section 7).

SNR shift ($+2.18$ dB mean). Weaker noise scales $\ell_t(k)$ uniformly, proportionally reducing $\Delta_{\min}$ and extending T_{mix}. The mild degradation on impulse conditions (-0.30 dB min) reflects this slower identification under reduced SNR. Structured conditions remain largely unaffected because the relative gap between experts is amplitude-invariant.

Fast regime switching ($+0.51$ dB mean). At 8 ms switching rates, the regime duration is comparable to T_{mix}, so the router never fully mixes before the next change. OFTRL degrades gracefully toward the bank-mean performance; all 13 conditions remain positive, confirming the router does not actively harm performance even when operating outside its ideal regime.

Loud real-world sources ($+5.64$ dB mean). Scaled DEMAND database noise ($1.5\times$ amplitude) tests in-distribution generalisation on real recordings. OFTRL maintains high NR because expert niches are amplitude-invariant: the routing signal $\ell_t(k)$ scales uniformly across all experts, so relative gaps $\Delta_{\min}$ and win-shares are preserved.

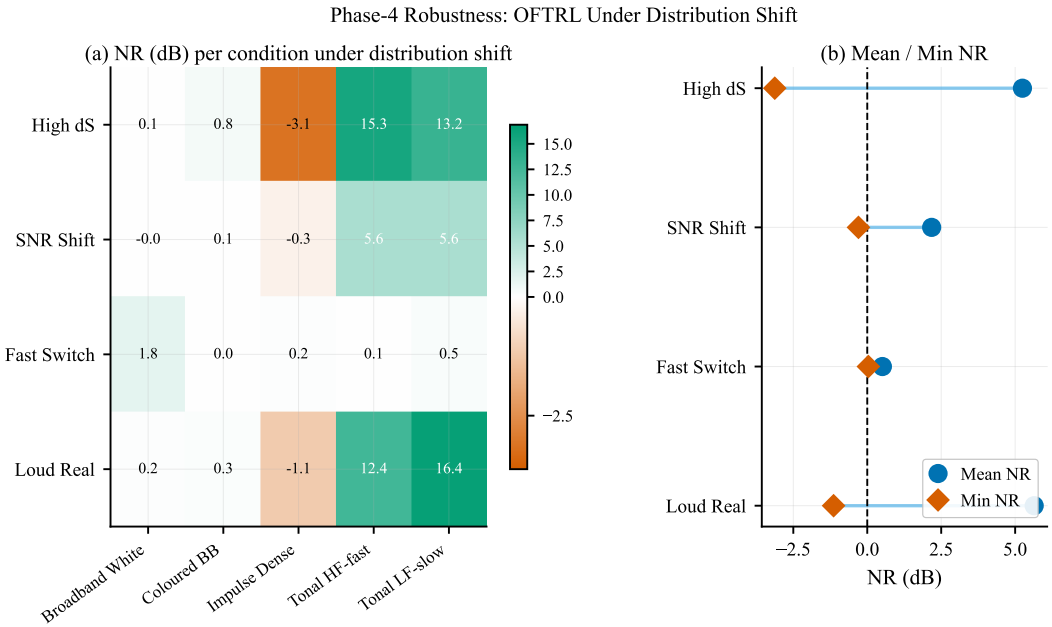

Figure 9: Per-condition NR under four perturbations vs. baseline OFTRL (grey). Mean NR positive throughout; single failure (`impulse_sparse`, high ΔS) is the documented design limitation (§7).

H Acoustic Noise Regime Descriptions

The 13 evaluation conditions span six stationary and seven non-stationary acoustic regimes, designed to cover the range of noise types encountered in real-world ANC deployments. Table 10 describes each regime's key spectral and temporal properties, its dominant expert (highest win-share from Table 2), and its empirical $\Delta_{\min}$.

Stationary regimes test steady-state routing accuracy under fixed noise characteristics. Non-stationary regimes test the router's ability to detect and adapt to abrupt or gradual changes in the optimal expert. The conditions were generated synthetically or drawn from the DEMAND database [33] (available at `https://zenodo.org/record/1227121` and on Kaggle): synthetic conditions permit exact control over temporal structure (switching rate, impulse rate, amplitude modulation frequency); DEMAND-derived conditions provide realistic spectral complexity and room acoustics.

Table 10: Acoustic noise regimes. S = stationary, NS = non-stationary. Dom. expert: highest win-share (Table 2). Src: Syn. = synthetic, D = DEMAND [33].

Condition	**Cls.**	**S/NS**	**Key properties**	**Dom. expert**	**Src**	$\Delta_{\min}$
Stationary						
Tonal LF-slow	Tonal	S	Pure tone 80–300 Hz; amp. drift <0.5 Hz	E2 (RLS λ=.9998)	Syn.	47.2
Tonal HF-fast	Tonal	S	Pure tone 1–4 kHz; AM at 5–20 Hz	E1 (RLS λ=.9990)	Syn.	38.9
Coloured BB	Bband	S	Pink (-3 dB/oct) 20 Hz–8 kHz	E3 (BLMS B=32)	D	8.3
HF hiss	Bband	S	White noise high-passed at 2 kHz	E2 (RLS λ=.9998)	Syn.	1.2
Wideband struct.	Tonal	S	3–5 harmonics + broadband floor	E2 (RLS λ=.9998)	Syn.	51.4
Broadband white	Bband	S	Flat 20 Hz–8 kHz (iid Gaussian)	E3 (BLMS B=32)	Syn.	0.4
Non-stationary						
Impulse sparse	Imp.	NS	<10 imp./s on BB background	E0 (FxNLMS L=256)	Syn.	22.7
Impulse dense	Imp.	NS	50–200 imp./s + aftershock	E4 (BLMS B=64)	Syn.	4.1
Cyclostationary	Cycl.	NS	AM broadband; 120 Hz modulation	E3 (BLMS B=32)	Syn.	0.9
Nonlinear cplx.	Cplx.	NS	Gliding harmonics + nonlinear interaction	E3 (BLMS B=32)	Syn.	2.3
Switch 250 ms	Swch.	NS	Tonal↔BB at 250 ms	E2 (RLS λ=.9998)	Syn.	5.8
Switch fast	Swch.	NS	Tonal↔BB at 8 ms	E1 (RLS λ=.9990)	Syn.	7.1
Mixed even	Mixed	NS	Stat. (25 ms) + NS transitions	E3 (BLMS B=32)	D	3.4

Regime design rationale. The five condition classes are chosen to span the key dimensions of expert-bank heterogeneity:

- *Tonal / wideband structured.* Large $\Delta_{\min}$ ($\approx$ 39–51, $\hat{T}_{\text{mix}} \leq 12$ steps): RLS experts (E1, E2) achieve near-optimal steady-state error through frequency-selective tracking that FxNLMS cannot match at the same filter length. These are the primary beneficiaries of fast routing and the conditions where OFTRL most strongly outperforms Hedge (Table 7).
- *Broadband / white.* Very small $\Delta_{\min}$ (< 10, $\hat{T}_{\text{mix}} > 50$ steps): No algorithm has a structural advantage on spectrally flat noise; win-shares are nearly uniform across E0–E4. These conditions

verify that OFTRL does not actively harm performance when mixing is slow—and that context cannot help either (gain is at most $O(\epsilon^2)$ per Theorem 2).

- *Impulsive.* Contrasting $\Delta_{\min}$: sparse (22.7, $\hat{T_{\text{mix}}} = 20$) vs. dense (4.1, $\hat{T_{\text{mix}}} = 106$). Sparse impulses expose E0's advantage (high-resolution FIR recovers quickly from brief transients); dense impulses favour E4 (block gradient amortises per-impulse noise). The contrast validates that $\Delta_{\min}$ correctly predicts which conditions benefit from fast routing.
- *Cyclostationary / nonlinear.* Low $\Delta_{\min}$ (< 2.4): no single expert dominates the quasi-periodic non-stationarity; these serve as null-hypothesis conditions for context injection, where theory predicts and experiment confirms no routing algorithm offers substantial improvement.
- *Switching / mixed.* Moderate $\Delta_{\min}$ (5.8–7.1, $\hat{T_{\text{mix}}} = 61$–$75$) with an abrupt change structure that creates the oracle hard-switch penalty. OFTRL's soft weight redistribution eliminates this penalty, achieving $+4.02$ dB vs. oracle $+2.13$ dB on Switch 250 ms-the strongest example of OFTRL's super-oracle advantage in the results.

Context utility taxonomy. Table 11 extends the analysis to a broader taxonomy of routing settings derived from Theorem 3 and Definition 2, identifying when loss-dominance holds, fails, or is ambiguous.

Table 11: Context utility in adaptive expert routing. Derived from Theorem 3 and Definition 2. Our ANC system falls in row 1.

Setting	$\Delta_{\min}$	T_{mix}	**Feedback**	**Context utility**
High-gap stationary *(this work)*	Large (>20)	Fast (<20)	Full-info	**Harmful.** Loss resolves arm identity before context stabilises; bias accumulates post-mixing.
Low-gap stationary	Small (<5)	Slow (>200)	Full-info	**Negligible.** No routing method gains substantially; $\Delta_{\min}$ too small for any routing advantage.
Switching / non-stat.	Moderate	Moderate	Full-info	**Marginal.** Context may aid switch detection; γ-discounting handles this without context here.
Bandit (partial)	Any	N/A	Partial	**Potentially helpful.** Context provides exploration signal unavailable from arm losses alone.
Slow adaptation	Any	Large	Full-info	**Potentially helpful** if context lag $<$ mixing time.