PHYSICAL ADAPTIVE CONTROL BENCHMARK

FARO vs RLS-CE

Online adaptation on a physical RC plant after regime mutation

Result: FARO recorded lower aggregate IAE than an online RLS certainty-equivalent controller under information-symmetric physical testing. FARO won 3 of 4 scored profiles; the advantage was negligible in FAST and expanded materially in SLOW.

150,859.68	155,106.36	2.74%	3 / 4
FARO aggregate IAE	RLS-CE aggregate IAE	lower aggregate IAE for FARO	profiles won by FARO

0.09%	4.29%	17.65%	PASS
FAST aggregate FARO edge	SLOW aggregate FARO edge	lower mean tail RMSE for FARO	timing comparability

What this test asks. Can FARO match or outperform a controller that explicitly identifies a first-order plant online using Recursive Least Squares and then applies certainty-equivalent one-step control, without FARO receiving a plant model, future FAST/SLOW characterization, or post-mutation tuning?

Answer within this protocol: yes on aggregate. The result supports a physical advantage in this tested plant and protocol; it does not establish universal superiority over RLS, adaptive control, or model-based control.

Information-symmetric adaptive comparison

RLS is an estimator, not a controller. The baseline is therefore explicitly defined as RLS-CE: Recursive Least Squares identification of a normalized first-order ARX model plus certainty-equivalent one-step control. Both arms use the same physical observations and the same +/-12 DAC action limit.

Dimension	FARO	RLS-CE
Preparation	384 MEDIUM decisions	384 MEDIUM decisions
Starting point for score	Prepared MEDIUM snapshot	Prepared MEDIUM snapshot
After mutation	Adapts online	Adapts online
Plant tau / explicit physical model	Not provided	Not provided a priori
Future FAST/SLOW profiles	Not provided	Not provided
Post-mutation tuning	None	None
Action limit	+/-12 DAC counts	+/-12 DAC counts
Primary metric	Raw nominal IAE	Raw nominal IAE

RLS-CE preregistration

Parameter	Value
Estimator	RLS on x_next = a*x + b*v + c
Normalization	x=(ADC-2048)/2048; v=(DAC-2048)/2048
Forgetting factor (lambda)	0.995
Initial covariance	P0 = 1000 I
Minimum transitions before model-based control	16
Bootstrap behavior	Only allowed +/-12 direction toward reference

Parameter	Value
Hyperparameter search on physical score	None

The original software QA was run before the physical experiment. RLS-CE numerically won both simulated batteries: NOMINAL 100,794.96 vs FARO 100,938.00; SHIFTED 101,212.32 vs FARO 101,609.04. The physical experiment therefore did not begin from a simulation that already favored FARO.

MEDIUM preparation -> FAST score -> SLOW score

Stage	Selector	Purpose	Scored?
MEDIUM	A21	Prepare FARO and RLS-CE: 384 decisions each	No
FAST	A18	FAST_UP + FAST_DOWN, 540 decisions per arm/profile	Yes
SLOW	A24	SLOW_UP + SLOW_DOWN, 540 decisions per arm/profile	Yes

Electrical qualification

Each regime passed the existing ZS1 Q gate before control evaluation. MEDIUM and FAST used the original center conditioning. For SLOW only, the legacy absolute 2048 +/-16 center condition was replaced before scoring by a direction-controlled 512 -> 2048 stabilization using the Q acceptance policy: |PA5-command| <=128, last-10 span <=8, minimum 30 observations, maximum 30 seconds per phase. The same conditioning was applied to FARO and RLS-CE.

Regime	Q 512	Q 2048 #1	Q 3584	Q 2048 #2	Gate
MEDIUM	549	2032	3535	2052	PASS
FAST	529	2038	3559	2056	PASS
SLOW	632	1925	3460	2169	PASS

Protocol integrity: no scored run was excluded, no performance retry was used, controller snapshots were restored for each scored profile, and the order was alternated across profiles.

Important scope note: controllers ran on the host while the RC plant and ADC/DAC path were physical. This benchmark does not claim native MCU controller latency. The result files explicitly mark exact_hardware_adc_timestamps=false.

FARO won 3 of 4 physical profiles

Profile	Order	FARO IAE	RLS-CE IAE	Winner	FARO rel. edge	Timing
FAST_UP	RLS-CE -> FARO	28,773.24	28,721.40	RLS-CE	-0.18%	PASS
FAST_DOWN	FARO -> RLS-CE	28,452.48	28,557.24	FARO	+0.37%	PASS
SLOW_UP	FARO -> RLS-CE	48,051.36	49,547.88	FARO	+3.02%	PASS
SLOW_DOWN	RLS-CE -> FARO	45,582.60	48,279.84	FARO	+5.59%	PASS

RLS-CE	FARO	FARO	FARO
FAST_UP winner (-0.18% FARO edge)	FAST_DOWN winner (+0.37%)	SLOW_UP winner (+3.02%)	SLOW_DOWN winner (+5.59%)

FAST_UP was the only profile won by RLS-CE, and the margin was small: 51.84 count-seconds. The separation increased after transition to the SLOW regime, where FARO recorded 1,496.52 less IAE in SLOW_UP and 2,697.24 less in SLOW_DOWN.

Because only four scored profiles were run, profile-win counts are descriptive. No formal statistical significance claim is made from 3/4 alone.

FAST: practical parity. SLOW: clear separation.

57,225.72	57,278.64	0.09%	1 / 1
FAST FARO IAE	FAST RLS-CE IAE	FAST lower IAE for FARO	FAST profile split

FAST is effectively parity: FARO is lower by only 52.92 count-seconds in the two-profile aggregate, a 0.09% relative difference. One FAST profile went to each controller.

93,633.96	97,827.72	4.29%	2 / 2
SLOW FARO IAE	SLOW RLS-CE IAE	SLOW lower IAE for FARO	SLOW profiles won by FARO

SLOW is where the physical difference appears: FARO reduced aggregate IAE by 4,193.76 count-seconds (4.29%) and won both mirror profiles. The result is consistent with a stronger FARO response to this larger physical regime shift, but it should not be generalized beyond the tested plant and protocol.

Overall: FARO 150,859.68 vs RLS-CE 155,106.36. Difference 4,246.68 count-seconds; FARO relative reduction 2.74%; symmetric FARO advantage 2.78%.

Accuracy improved with a small actuator-movement cost

Metric	FARO	RLS-CE	Reading
Total DAC variation	25,200	24,789	FARO +1.66%
Mean tail RMSE	84.710	102.868	FARO 17.65% lower
Host time-weighted post-error proxy	152,485.43	156,318.46	FARO 2.45% lower
Cycles >150 ms	0	1	RLS-CE event only in SLOW_DOWN
Timing comparability	PASS	PASS	All four pairs comparable

Tail RMSE by profile

Profile	FARO	RLS-CE	Lower
FAST_UP	1.816	1.943	FARO
FAST_DOWN	1.762	2.723	FARO
SLOW_UP	175.190	201.545	FARO
SLOW_DOWN	160.071	205.259	FARO

The trade-off is explicit: FARO used 411 additional DAC counts of total variation across all scored profiles (1.66% more) while reducing aggregate IAE by 2.74% and mean tail RMSE by 17.65%. This is not a free improvement; it is a measured accuracy-versus-actuation trade-off.

06 / INTERPRETATION

What the result supports - and what it does not

Within this physical protocol, FARO achieved lower aggregate tracking error than an online RLS certainty-equivalent controller without receiving an explicit plant model or post-mutation tuning.

Supported

1. FARO achieved 2.74% lower aggregate IAE over four physical scored profiles.
2. FAST was effectively parity, while FARO's advantage expanded to 4.29% aggregate IAE in SLOW.
3. FARO won 3/4 profiles and had 17.65% lower mean tail RMSE.
4. The RLS-CE baseline was adaptive online and had the same observation stream and action limit.
5. The pre-experiment simulations favored RLS-CE, so the physical winner was not forced by the software QA.

Not supported

1. No claim that FARO universally outperforms RLS, adaptive control, MPC, LQR, or model-based control.
2. No claim of statistical significance from four scored profiles.
3. No claim of native MCU controller latency or energy advantage from this experiment.
4. No claim that RLS itself is a controller; the comparator is specifically RLS + certainty-equivalent control.
5. No claim that the tested first-order ARX baseline exhausts all possible RLS-based controller designs.

Recommended benchmark wording: “Under information-symmetric physical testing, FARO achieved 2.74% lower aggregate IAE than an online RLS certainty-equivalent controller, winning 3 of 4 scored profiles. The difference was negligible in FAST dynamics and expanded to 4.29% aggregate IAE reduction in the SLOW regime.”

Raw trace identifiers

Profile	FARO trace SHA256	RLS-CE trace SHA256
FAST_UP	38d0a9512dd4eabe5a1824d7...	0c316caa849f4893e85c89a1...
FAST_DOWN	61029d7a4b6f0f4a72f5e909...	28a837c0272d8a8ce8e01ea2...
SLOW_UP	24cbfd1ba54a4aedfaec8684...	8e02abef54025012641b0212...
SLOW_DOWN	67bbf18b01760715aa1bab0e...	18994678c2c5f4335207ac7e...

Execution record

Item	Value
Execution record ID	FISICO_20261030_193029_783042
Firmware	Existing OSH_RC_BENCH_ZS1; no flash during benchmark
Engine dependency	FARO_PID_PLANTA_MUTANTE_ZS2B.py
Engine SHA256	8afeb3b375db92d281c60b29c5c4e0907080c1b60ff81fcdfa3a2b93dfe2288f
Scored profiles	4
Scored runs	8
Decisions per scored run	540
Scored controller decisions	4,320
Performance retries	0
Primary metric	0.12 * sum(abs(reference - raw post median))
Timing policy	<=5% cycles >150 ms and pair duration ratio <=1.10

Each completed trace was replay-audited by the benchmark script for controller decision consistency. Hardware truth was not revalidated by the replay itself. Full raw JSONL traces and state files remain the authoritative reproducibility artifacts.

Bottom line: this experiment closes a different credibility gap than the PID Oracle test. The Oracle test asks whether FARO survives a privileged classical baseline; this test asks whether FARO survives an information-symmetric online model-identification baseline.