

Online Resource 1

Stateful Runtime Adaptation under Workload and Thermal Change: An Audited Evaluation on a Heterogeneous SoC

Antonio Souto Rodriguez and José Reginaldo Hughes Carvalho

Institute of Computing, Federal University of Amazonas, Manaus, Brazil

Supplementary analysis, 26 October 2026

The Journal of Supercomputing

Corresponding author: antonio.rodriguez@icomp.ufam.edu.br

S1 Scope and precise estimands

This supplement reports the completed 18-mission comparison: six ERL dynamic missions and twelve static-policy missions, totaling 107.9996 hours. Each mission contains twelve scheduled workload phases and eleven replacements. Four workloads recur for three visits. The primary measurements are reported throughput, processing latency, sampled power and CPU temperature. Whole-mission host-timed meter integration is a separate outcome.

The numerical sources are the preserved, raw-derived analysis dated 23 October 2026, the final summary recalculation and the conservative forecast reanalysis dated 25 October. The current recheck verified the preserved package and reproduced surviving raw subsets; it did not reopen every original stream from all 18 missions. Section S11 specifies the verification boundary.

For policy a , profile p , run r , workload w and measured channel k , the primary within-run statistic and two-run centre are

$$m_{a,p,r,w,k} = \text{median}\{x_{t,k} : t \text{ belongs to workload } w \text{ in run } r\}, \quad (\text{S1})$$

$$\bar{m}_{a,p,w,k} = \frac{1}{2}(m_{a,p,1,w,k} + m_{a,p,2,w,k}). \quad (\text{S2})$$

The first expression pools the three workload visits *within one run*; it is not the median of three phase medians. Samples are never pooled across repetitions for the cell estimate. Both run values and their minimum–maximum range are retained in the CSVs. Relative comparisons are $100(\bar{m}_{\text{ERL}}/\bar{m}_a - 1)$. Selecting the largest of the three static throughput centres identifies the best *observed supplied policy*, not an optimization oracle.

Repetition identifiers are chronological. They are not randomized matched blocks across policies. There are two independent missions per observed condition. The many frame, phase and controller-period records do not increase that independent replication count. Figures display individual runs and observed ranges, without population confidence intervals or frame-level significance tests.

Table S1: Measurement interpretation retained throughout this supplement.

Quantity	Interpretation and limit
FPS	Sampled reported throughput estimator; equal numerical output is unverified.
Processing ms	Processing interval; not sensor-to-output latency.
Trace power W	Median of reported power samples belonging to a workload.
CPU temperature	Recorded CPU temperature; not the supervisor’s two-sensor maximum.
Mission energy	Channel-0 meter power integrated on host time over the manifest window.
GPU enabled	Permission/configuration; not per-frame proof of GPU execution.
C1a/C1b	Software policy/readback consistency; not physical clock residency.

S2 Campaign inventory and record counts

The timestamp identifies each mission uniquely. Dynamic identifiers have the form `pilot_THERMAL_{profile}_{timestamp}`. Static identifiers have the form `pilot_{arm}_THERMAL_{profile}_{timestamp}`. The complete strings and manifest declarations are in `authoritative_18run_inventory.csv`. In tables and figures, CPU, GPU and Balanced abbreviate CPU_STATIC, GPU_STATIC and BALANCED. Hist, Sobel, Gaussian and Median denote HistogramEqualization, SobelEdge, HeterogeneousGaussianBlur and MedianFilter.

Table S2: All supplied missions. Every row contains twelve phases and eleven workload replacements. Trace snapshots, retained metric rows and controller actions are different record types.

Profile	Policy	Rep.	Mission timestamp	Trace rows	Metric rows	Actions
Inactive	ERL	1	20261012_072555	42,550	466,897	26,907
Inactive	ERL	2	20261017_085639	42,519	435,813	26,972
Inactive	CPU	1	20261017_225210	42,538	428,607	–
Inactive	CPU	2	20261022_155716	42,539	428,916	–
Inactive	GPU	1	20261019_161507	42,648	513,850	–
Inactive	GPU	2	20261021_163625	42,636	512,843	–
Inactive	Balanced	1	20261018_153703	42,636	505,006	–
Inactive	Balanced	2	20261020_212425	42,615	500,514	–
Moderate	ERL	1	20261011_093117	42,566	463,557	26,976
Moderate	ERL	2	20261016_150925	42,530	453,997	26,972
Aggressive	ERL	1	20261010_203819	42,461	369,096	26,951
Aggressive	ERL	2	20261016_074913	42,494	396,069	26,945
Aggressive	CPU	1	20261019_073140	42,514	452,040	–
Aggressive	CPU	2	20261023_074052	42,545	403,033	–
Aggressive	GPU	1	20261020_075035	42,601	446,425	–
Aggressive	GPU	2	20261022_074229	42,642	465,245	–
Aggressive	Balanced	1	20261018_091202	42,609	422,997	–
Aggressive	Balanced	2	20261021_073557	42,607	412,225	–

There are 766,250 trace snapshots and 8,077,130 retained metric rows across the full campaign. The six dynamic missions contain 161,723 action records and 2,585,429 retained metric rows. These totals must not be described interchangeably as “control periods.” The 72 run–workload summaries correspond to 18 missions times four workloads. The 432 phase-summary rows contain both the 216 full phases and their first-30-seconds-excluded sensitivities.

The application-level caution/warning/critical thresholds are 75/85/95 °C for inactive, 55/62/70 °C for moderate and 45/50/55 °C for aggressive. These are policy settings, not target temperatures or guarantees. Only ERL has moderate-profile missions; no moderate static comparison is imputed.

CPU_STATIC requests four CPU workers without GPU; GPU_STATIC requests one worker with full GPU split; BALANCED requests two workers with half split. The static launcher applies its own WARNING-stage GPU withdrawal gate. It does not execute the entire dynamic governor/AWM/GA path. Shared profile names therefore do not make the complete policy bundles identical interventions.

S3 Complete policy–profile–workload centres

Table S3: Arithmetic means of two run-level medians for every observed cell. Each run median pools three workload visits. No cell is filled from a missing moderate static arm. All full-precision run values and observed ranges accompany this supplement in the CSVs.

Workload	Policy	FPS	Processing (ms)	Power (W)	CPU (°C)
Inactive					
Hist	ERL	44.90	1.125	6.338	56.00
Hist	CPU	46.45	1.119	6.349	57.25
Hist	GPU	46.05	3.313	6.466	57.50
Hist	Balanced	46.24	3.316	6.498	58.50
Sobel	ERL	44.97	7.951	6.393	58.00
Sobel	CPU	45.19	8.628	6.372	57.25
Sobel	GPU	45.74	2.254	6.464	57.25
Sobel	Balanced	45.56	2.250	6.488	58.00
Gaussian	ERL	37.18	14.839	6.650	58.50
Gaussian	CPU	35.16	23.146	6.564	58.50
Gaussian	GPU	45.90	5.582	6.527	57.50
Gaussian	Balanced	42.26	5.767	6.596	58.75
Median	ERL	33.95	11.680	6.376	57.50
Median	CPU	30.06	28.036	6.715	59.25
Median	GPU	43.52	10.744	6.842	59.00
Median	Balanced	42.75	11.140	6.879	60.00
Moderate					
Hist	ERL	41.84	1.146	6.196	55.50
Sobel	ERL	40.24	2.048	5.986	55.50
Gaussian	ERL	36.83	6.564	6.585	56.75
Median	ERL	37.83	8.788	6.479	57.75
Aggressive					
Hist	ERL	41.14	1.285	5.585	52.25
Hist	CPU	46.86	1.119	6.145	53.25
Hist	GPU	48.20	1.161	6.223	52.75
Hist	Balanced	46.29	1.131	6.393	56.00
Sobel	ERL	37.99	9.749	5.183	52.25
Sobel	CPU	46.33	6.376	6.128	53.00
Sobel	GPU	47.42	1.071	6.221	52.50
Sobel	Balanced	44.94	7.725	6.389	56.00
Gaussian	ERL	31.24	26.051	5.626	53.00
Gaussian	CPU	38.44	18.047	6.389	54.75
Gaussian	GPU	42.89	12.404	6.418	53.75
Gaussian	Balanced	37.35	19.983	6.592	56.75
Median	ERL	20.73	40.031	5.512	53.50
Median	CPU	26.27	33.800	6.498	55.50
Median	GPU	17.31	54.199	6.438	54.25
Median	Balanced	16.87	54.465	6.570	57.25

These values characterize the recorded implementations. CPU/GPU output equivalence, identical live inputs and matching thermal histories have not been established. Cross-policy latency differences must not be interpreted as speedups for demonstrated equivalent work.

S4 Individual-run performance and environmental measurements

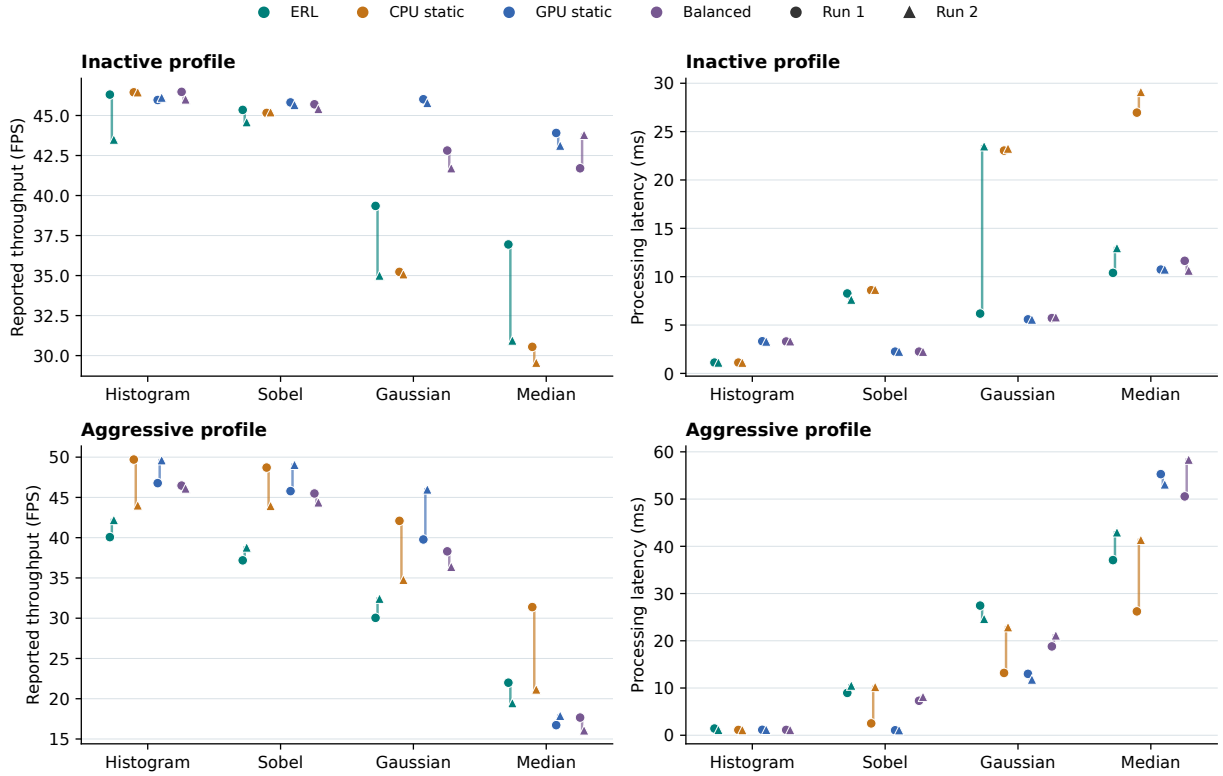


Figure S1: Primary run-level throughput and processing-latency medians. Each policy has two individual mission markers per workload. Marker shape identifies chronological repetition; line segments connect that policy’s observed minimum and maximum and are not confidence intervals or paired causal effects. Axis scales differ by measured channel.

The largest observed static throughput centre exceeds ERL in all eight profile/workload comparisons. Inactive-profile deficits are 3.3%, 1.7%, 19.0% and 22.0% for Hist, Sobel, Gaussian and Median, respectively. Aggressive-profile deficits are 14.7%, 19.9%, 27.2% and 21.1%. These comparisons retain both repetitions of every supplied static arm, including aggressive CPU_STATIC’s unusually fast first recording.

A narrower favorable case occurs for aggressive Median: ERL’s centre is 20.731 FPS, compared with 17.306 for GPU_STATIC and 16.872 for BALANCED, gains of 19.8% and 22.9%. CPU_STATIC remains higher at 26.272 FPS. The corresponding ERL processing centre is 40.031 ms, versus 54.199 and 54.465 ms for GPU_STATIC and BALANCED. Without equal-output evidence, these observations describe the measured implementation paths rather than equivalent-work speedups.

Power and CPU temperature at workload scale

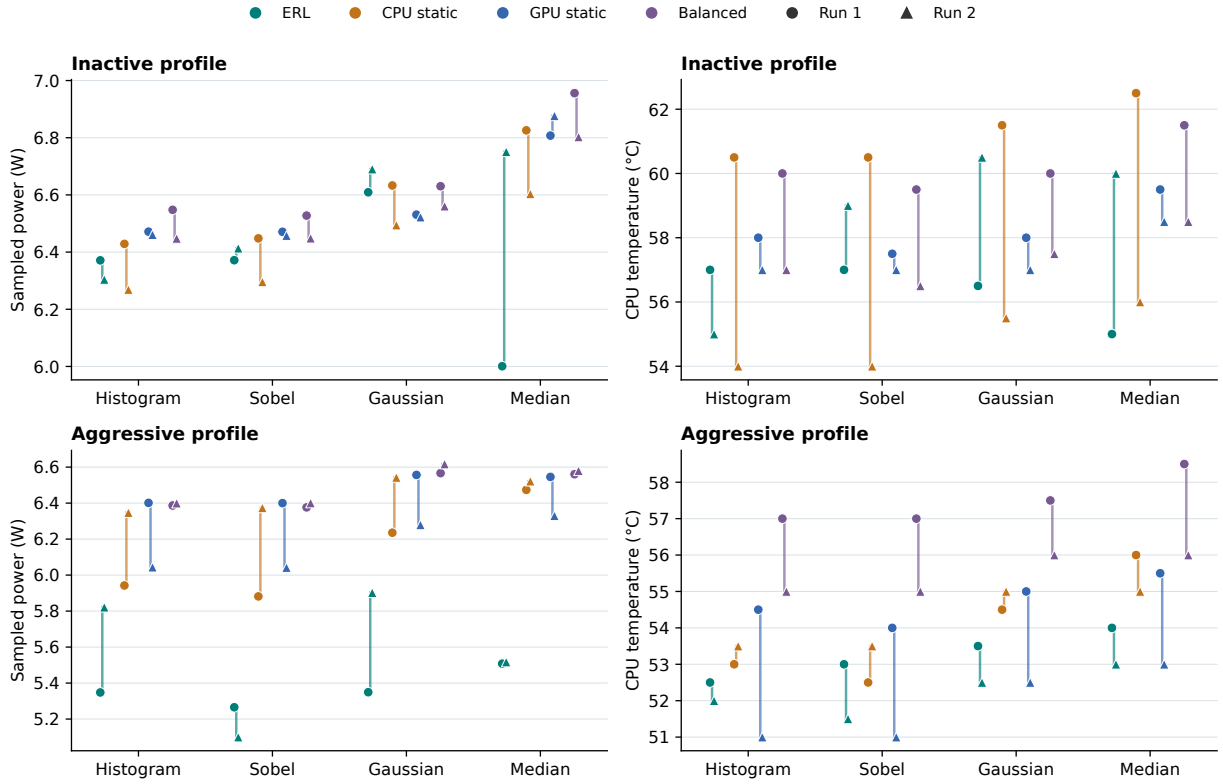


Figure S2: Primary run-level trace-power and CPU-temperature medians. The lower aggressive-profile ERL power observations coexist with lower throughput for most comparisons. These panels do not imply equal completed work, equal thermal histories or equivalent policy interventions. Moderate-profile ERL values are included in Table S3; no moderate static observations exist.

Under the aggressive profile, each ERL workload's sampled-power centre is 0.560–1.206 W below each supplied static comparator. The direction holds for all four cross-run value combinations in each comparison. This is an observed ordering over two runs per policy; four value combinations are not four independent experiments.

Lower CPU temperature also does not establish thermal safety or threshold compliance. The application supervisor uses the maximum of CPU and GPU sensor values, while this panel displays only CPU temperature. Profiles are configured escalation thresholds. Time above a threshold and adherence to the GPU-denial predicate answer different questions, and neither can be inferred solely from a workload median.

S5 Host-time meter integration and interval sensitivity

Channel-0 power is calculated as voltage times current and trapezoidally integrated on `system_time_us`, clipped to the manifest mission window with interpolated endpoints. The full integration linearly interpolates gaps. The sensitivity removes intervals longer than three seconds and divides retained energy by retained interval duration. Thus its mean-power change is not the energy of an identical-duration mission.

Table S4: Mission energy (kJ), integrated mean power (W), largest host interval (s), number of intervals over 3s, supported mission coverage (%) and supported-minus-full mean power (W).

Profile	Policy	Rep.	Energy	Mean W	Max gap	Gaps	Support	ΔW
Inactive	ERL	1	133.831	6.1959	2.325	0	100.0000	-0.000000
Inactive	ERL	2	136.211	6.3061	2.412	0	100.0000	-0.000000
Inactive	CPU	1	141.344	6.5437	7.173	2	99.9338	+0.000529
Inactive	CPU	2	137.802	6.3797	7.143	3	99.9009	+0.001133
Inactive	GPU	1	140.625	6.5105	1.996	0	100.0000	+0.000000
Inactive	GPU	2	140.919	6.5241	2.023	0	100.0000	+0.000000
Inactive	Balanced	1	142.880	6.6148	7.124	1	99.9670	+0.000795
Inactive	Balanced	2	140.776	6.5174	2.084	0	100.0000	+0.000000
Moderate	ERL	1	133.920	6.2000	8.172	5	99.8293	+0.000984
Moderate	ERL	2	131.087	6.0689	2.308	0	100.0000	-0.000000
Aggressive	ERL	1	115.673	5.3552	7.216	6	99.8570	+0.000516
Aggressive	ERL	2	120.274	5.5683	2.544	0	100.0000	-0.000000
Aggressive	CPU	1	131.412	6.0839	7.118	3	99.9013	+0.001236
Aggressive	CPU	2	138.462	6.4103	8.198	5	99.8291	+0.001060
Aggressive	GPU	1	138.781	6.4251	7.136	2	99.9339	+0.000521
Aggressive	GPU	2	132.033	6.1126	7.102	2	99.9343	+0.000838
Aggressive	Balanced	1	138.453	6.4099	8.119	5	99.8303	+0.001153
Aggressive	Balanced	2	139.266	6.4475	7.115	3	99.9012	+0.001449

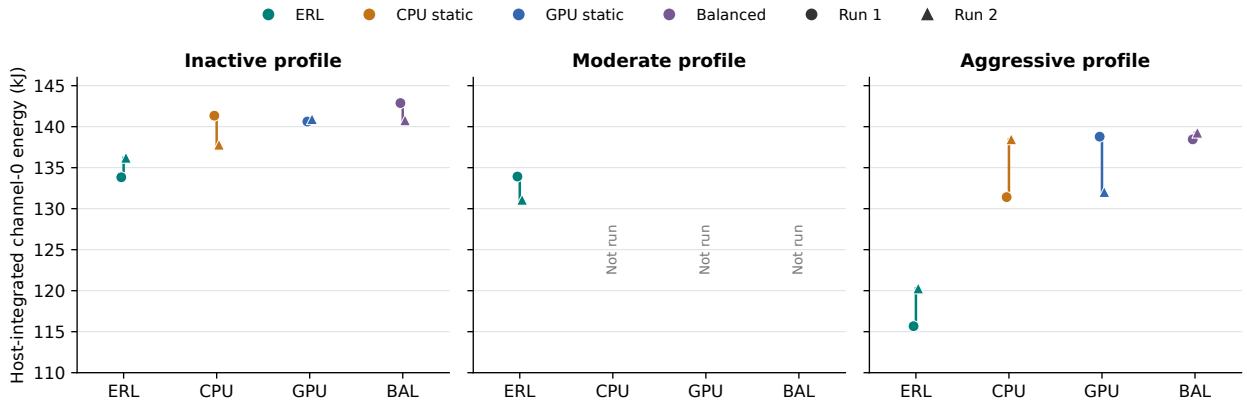


Figure S3: Individual approximately six-hour mission energies. Missing moderate static conditions are explicitly unobserved. Energy is a mission-total quantity, not energy per delivered equivalent output.

The largest host gap is 8.197825 s. Supported coverage is at least 99.8291%, and removing intervals over three seconds changes mean power by at most 0.001449 W. Device-clock nonpositive steps and a reported jump of approximately 10,721 s are distinct from these host intervals; they are not treated as missing acquisition time. These checks establish robustness to short host gaps, not meter rail attribution, calibration or sensor freshness.

S6 Strict four-metric reference score and weights

The score is a supplementary descriptive normalization. For each profile/workload, let b_k and u_k be the minimum and maximum of the *three static two-run centres* on channel k , and let $d_k = u_k - b_k$. The fixed cost coordinates are

$$z_F(x) = (u_F - x_F)/d_F, \quad z_k(x) = (x_k - b_k)/d_k \quad (k = L, P, T), \quad (\text{S3})$$

where F, L, P, T denote FPS, processing latency, power and CPU temperature. Values are not clipped: an ERL point can lie outside the observed static range. For nonnegative weights summing to one, $J(x) = \sum_k w_k z_k(x)$. With $J_{\min}$ and $J_{\max}$ taken over the three static policies,

$$G(x) = \frac{J(x) - J_{\min}}{J_{\max} - J_{\min}}. \quad (\text{S4})$$

All four static channel ranges must reach the fixed floors (2 FPS, 0.5 ms, 0.25 W, 1 °C), and the static composite range must be at least 0.05. Failure of any floor leaves the strict score undefined; it does not set a missing metric to zero or silently renormalize the remaining weights.

Table S5: Observed static-centre ranges and strict metric-floor admission. A dagger marks a range below its floor. Only aggressive Sobel passes all four floors.

Profile	Workload	FPS	ms	W	°C	All pass
Inactive	Hist	0.406 [†]	2.198	0.149 [†]	1.25	No
Inactive	Sobel	0.555 [†]	6.378	0.116 [†]	0.75 [†]	No
Inactive	Gaussian	10.744	17.564	0.069 [†]	1.25	No
Inactive	Median	13.461	17.292	0.165 [†]	1.00	No
Aggressive	Hist	1.909 [†]	0.042 [†]	0.249 [†]	3.25	No
Aggressive	Sobel	2.482	6.653	0.261	3.50	Yes
Aggressive	Gaussian	5.542	7.579	0.203 [†]	3.00	No
Aggressive	Median	9.400	20.664	0.132 [†]	3.00	No

Table S6: Aggressive Sobel strict reference gaps under four stated weight vectors. There are two defined dynamic run-workload scores out of sixteen candidates at the base floors, for each weight set.

Weight choice	(w_F, w_L, w_P, w_T)	Rep. 1	Rep. 2
Protocol	(0.5, 0.1, 0.2, 0.2)	1.591	1.038
Throughput	(0.7, 0.1, 0.1, 0.1)	2.753	2.196
Equal	(0.25, 0.25, 0.25, 0.25)	0.492	0.087
Power	(0.2, 0.1, 0.5, 0.2)	-1.046	-1.666

The weight choice reverses the apparent aggregate conclusion: the protocol weights give gaps 1.591 and 1.038, whereas power emphasis gives -1.046 and -1.666 . Neither a negative gap nor a value below one establishes global optimality, equal-output efficiency or a causal mechanism advantage. Halving every floor admits eight of sixteen candidate run-workload scores for each stated weight set; doubling every floor admits none. These are admission sensitivities, not evidence for choosing a favorable threshold after seeing the result.

S7 Post-hoc two-channel score and alternate summaries

The throughput–processing-latency score G_{FL} is a separate, explicitly post-hoc descriptive view. It uses the fixed static bounds from Section S6, retains the FPS and processing-latency floors, and renormalizes the corresponding protocol weights to $(5/6, 1/6)$. It does not replace an inadmissible four-metric result. For the inactive Gaussian and Median cells, the retained values are:

Table S7: Post-hoc throughput–processing-latency gap for inactive-profile missions. These selected two-channel summaries omit power and temperature.

Workload	Rep. 1	Rep. 2
Gaussian	0.513846	1.015132
Median	0.403514	0.799512

The reference is still only the three observed static policies. A gap of one denotes their worst composite score under these weights, not the worst achievable system configuration. A gap of zero denotes their best composite score, not a global optimum. Since useful output equivalence and common live input are unresolved, these normalizations describe measured policy bundles rather than an algorithm-equivalent speed–energy frontier.

Primary pooling versus median of phase medians

The median of the three full-phase medians is supplied as an alternate estimand for each run/workload. It weights visits equally rather than weighting every retained trace sample equally. It is never silently substituted into the primary tables. The largest absolute difference across the 72 run/workload summaries is 2.3083 FPS, 9.01675 ms, 0.637785 W and 1.5 °C respectively. The largest FPS difference occurs for inactive ERL repetition 2 Median (primary 30.94765 versus alternate 28.63935 FPS). The largest latency and power differences occur for aggressive ERL repetition 2 Median (42.98285 versus 33.96610 ms; 5.516935 versus 6.154720 W). This sensitivity demonstrates that aggregation choices matter when phases contain trends.

Removal of the first 30 seconds of each phase

The full 1,800-s phase is primary. Removing its first 30 seconds yields 216 alternate phase summaries. Across these phases, changes range from -0.3468 to $+1.3715$ FPS, -1.98335 to $+0.3462$ ms, -0.12717 to $+0.0960$ W and 0 to $+0.5$ °C. This is a measurement-window sensitivity, not a guarantee of steady state. The beginning of a phase can contain adaptation and is not automatically invalid.

Cap-consistent trace subset

The preserved trace-level sensitivity retains positive reported cap/current fields, command/readback agreement and current-state values not above the expected cap. It retains 242,007 of 255,120 dynamic trace snapshots, excluding 13,107 above-cap snapshots and six unset/nonpositive snapshots. These are trace snapshots; the separate action-level C1b count is 3,701 of 161,723 control periods. Sampling and admission units differ. Across 24 dynamic run/workload summaries, the largest absolute changes after this filter are 3.0358 FPS, 4.0065 ms and 0.30034 W. Filtering a controller-dependent state can change the mix of operating periods, so these values are sensitivities rather than corrected causal estimates. Physical clock residency remains unmeasured.

S8 Continuity, computation and within-run uncertainty

All 161,723 action records satisfy C1a software cap/readback equality. C1b records 3,701 reported-current-state values above the cap (2.2885%). All 66 dynamic workload replacements preserve the logged generation counter. This establishes counter continuity, not byte-identical population or surrogate contents. The median replacement downtime ranges from 10.9001 to 24.7196 ms across missions; the largest recorded downtime is 662.179 ms.

Table S8: Dynamic operational summaries. GA duty is the accumulated named GA-computation duration divided by elapsed mission time. It is not total controller/OS/telemetry overhead.

Profile	Rep.	Actions	C1b flags	Swap med. (ms)	GA med. (ms)	GA P95 (ms)	GA duty (%)
Inactive	1	26907	222	14.39	2.30	26.11	0.85
Inactive	2	26972	265	13.63	2.04	22.95	0.75
Moderate	1	26976	246	11.21	2.25	23.52	0.82
Moderate	2	26972	144	10.90	2.76	25.30	0.96
Aggressive	1	26951	2621	24.72	3.24	112.48	3.48
Aggressive	2	26945	203	19.01	2.44	77.59	1.40

Median named GA computation is 2.04359–3.24333 ms (0.255–0.405% of a nominal 800-ms period). The per-period 95th percentiles are 22.947645–112.477 ms (2.868–14.060%). Seven records exceed 800 ms across all six missions. This per-period statistic differs from accumulated GA duty (0.7525–3.4794%). It must not be described as one overrun in every mission or as complete controller cost.

The preserved per-phase intervals use 1,000 bootstrap replicates of nonoverlapping 120-s blocks, fixed seed 20261023. They describe variability within an observed mission, conditional on this block construction. They do not provide six or more independent missions, a between-policy confidence interval, or causal evidence of memory retention. The FPS dependence diagnostic uses one-second bins and records the first lag with autocorrelation below 0.1, up to 120 s. Eleven of 72 dynamic phases fail to reach that threshold by the cap.

Table S9: Dynamic phases not decorrelated by 120 s. Phase numbering is zero-based. Late minus early compares the last and first 300-s FPS medians; it diagnoses observed change, not a memory effect.

Profile	Rep.	Phase	Visit	Workload	ACF(120)	Late – early FPS
Aggressive	1	1	1	Sobel	0.678	29.440
Aggressive	2	0	1	Hist	0.173	-3.974
Inactive	1	3	1	Median	0.371	11.187
Inactive	1	7	2	Median	0.131	-0.983
Inactive	1	10	3	Gaussian	0.153	6.851
Inactive	2	2	1	Gaussian	0.184	15.042
Moderate	1	0	1	Hist	0.130	-10.293
Moderate	1	2	1	Gaussian	0.246	9.381
Moderate	1	5	2	Sobel	0.168	-5.292
Moderate	2	0	1	Hist	0.140	-6.288
Moderate	2	8	3	Hist	0.162	2.912

Long dependence and nonstationary trends can impair block-bootstrap coverage. Visually separated visit-specific intervals therefore cannot establish that recurrence caused an improvement. The two independent missions per condition and all individually displayed mission values are the primary uncertainty information.

S9 Conservative forecast admission and baseline construction

The forecast analysis uses the preserved event-derived `paired_scores` and `phase_events` projections. It does not reopen raw JSONL or establish that the inspected source is the exact historical binary. Frozen logged predictions are paired to recorded outcomes by collector query identity. All outcome windows are labeled assigned; settings-change state is unknown (code 2). Processing-latency prediction is excluded (code 3), because the historical prediction target is an aggregate end-to-end channel whereas the outcome scored by the collector is processing latency.

The original projections contain 161,717 scored pairs. Sixty-six have outcome timestamps at or after the exit of the phase owning the prediction, exactly one per dynamic workload switch. Two further moderate repetition 1 pairs have prediction/application timestamps before recorded workload entry. Excluding all 68 leaves 161,649 pairs. Each retained pair satisfies $\text{prediction} \leq \text{apply} \leq \text{outcome}$ and the available recorded phase bounds. The final phase has no exit in this projection; its terminal bound is left open rather than invented.

This exclusion is conservative. Source inspection indicates that GA learning credit is cleared at workload change while the collector’s prior ticket may persist. A first incoming-workload observation can therefore be associated with an outgoing ticket. The two pre-entry exclusions do not prove that the wrong workload executed: GA can resume before the launcher records entry. Structural ordering plus recorded phase bounds does not establish the membership of underlying source frames or sensor acquisition intervals.

For each retained query, the persistence baseline uses the most recent retained same-workload outcome available strictly before the recorded prediction timestamp. The exponential moving average (EMA) baseline uses the same clean, strictly prior observations with fixed $\alpha = 0.30$, initialized from the first available observation. Both histories persist across visits. Excluded boundary outcomes are also removed from history; their influence is not allowed to propagate into the EMA. Four cold-start observations per mission have no history, leaving 161,625 identical targets for the three-predictor comparison.

Table S10: Retained forecast pairs and common baseline-comparable targets after phase-boundary exclusions.

Profile	Rep.	Clean pairs	Common pairs	No history
Aggressive	1	26,939	26,935	4
Aggressive	2	26,933	26,929	4
Inactive	1	26,895	26,891	4
Inactive	2	26,960	26,956	4
Moderate	1	26,962	26,958	4
Moderate	2	26,960	26,956	4

The inspected source constructs issue time from the selection boundary rather than independently instrumenting prediction completion. If that boundary is missing, a fallback derives it from apply completion. Baseline chronology is therefore conditional on the recorded event times and on unresolved source-to-binary provenance. The analysis is retrospective on selected trajectories; baselines were not deployed controllers and do not evaluate unselected configurations.

S10 Forecast comparison, recurrence and sensitivity

Table S11: Median absolute throughput error (FPS) on identical non-cold-start targets. The lower ERL error is a descriptive selected-forecast result, not evidence of improved scheduling or a causal benefit from memory retention.

Profile	Rep.	Pairs	ERL	Persistence	EMA (0.30)
Aggressive	1	26,935	3.091	5.449	4.722
Aggressive	2	26,929	2.836	3.705	3.186
Inactive	1	26,891	3.221	3.889	3.363
Inactive	2	26,956	3.027	3.724	3.164
Moderate	1	26,958	3.806	4.759	3.944
Moderate	2	26,956	3.367	4.177	3.549

Logged ERL forecasts have lower median absolute throughput error than both simple retrospective baselines in each mission. ERL MedAE is 2.836–3.806 FPS, versus 3.705–5.449 FPS for persistence and 3.164–4.722 FPS for EMA. ERL and EMA predictions coincide within 10^{-6} in 7.8–49.0% of comparable pairs, consistent with related exponentially weighted mechanisms; the comparison is not between wholly unrelated model families.

Table S12: Within-mission visit-specific ERL median absolute FPS error on the retained event-pair set. Only aggressive repetition 2 has lower visit-3 than visit-1 MedAE. Visit number is not a randomized intervention.

Profile	Rep.	Visit 1	Visit 2	Visit 3	V3 – V1
Aggressive	1	2.733	3.315	3.204	0.470
Aggressive	2	2.769	3.215	2.561	-0.208
Inactive	1	3.063	3.249	3.359	0.296
Inactive	2	2.587	3.153	3.392	0.805
Moderate	1	3.640	3.994	3.779	0.139
Moderate	2	3.292	3.436	3.378	0.086

There is no consistent recurrence-associated error reduction. At visits 2 and 3, recorded exact-bucket source status covers 99.54–99.88% and 99.69–100% of retained pairs, respectively. This describes a recorded prediction-source label. Exact-bucket support is shared across targets in the reference source and is not proof of a fresh, measured, independently calibrated observation for every target.

Requiring prediction timestamps to occur at least 1, 5 or 10 seconds after recorded phase entry changes each mission’s retained-set throughput MedAE by at most 0.0107 FPS. This additional phase-margin check is descriptive and does not prove that actuation has settled or that a sensor aggregate belongs solely to the assigned action.

Power and temperature error diagnostics are available in the machine-readable forecast outputs, but freshness, smoothing and attribution remain unresolved. They are not primary claims here. Processing-latency predictive accuracy is not claimed because its target is incompatible with the scored historical channel.

The result supported by this analysis is narrowly stated: selected ERL throughput forecasts predict the collector’s recorded target more accurately, in median absolute error, than these two fixed, strictly prior-history baselines across these six missions. It does not establish memory-retention benefit, positive transfer, counterfactual candidate accuracy, improved control decisions or causal performance gains. Two independent missions per profile remain the units of replication.

S11 Provenance, reproducibility and interpretation boundary

The final verification checked 276 preserved package members against their manifest hashes without failures. All 38 surviving local raw inputs matched their preserved SHA-256 inventory entries. Recomputing six static trace files produced 24 run/workload rows that exactly matched the preserved primary FPS, processing-latency, power and CPU-temperature medians. Recomputing the two inactive action files reproduced 53,879 actions, 53,879 C1a equalities and 487 C1b exceptions. These checks validate the surviving subset and the integrity of the preserved analysis; they do not constitute a new full raw rebuild of the 108-hour campaign.

Reproducing summary calculations from preserved intermediate records is distinct from regenerating those records from every original raw stream. The archive checksum checks and raw-subset recalculations above do not certify a complete independent raw replay. The accompanying raw-data analysis entry point is provided for execution with the complete corpus; this supplement does not assert a new full-corpus raw replay result. The complete eighteen-mission raw corpus, its analysis driver, the full-corpus replay log and the per-file checksums are version 2 of Zenodo record 10.5281/zenodo.23350634, which is published as <https://doi.org/10.5281/zenodo.23676764>.

The public source redistribution contains 90 unchanged files from the reviewed 91-file snapshot. The omitted file is `Stage_01/Others/LKof_defines.h`, which carries a proprietary notice. The original per-file hash manifest is retained for correspondence; the public distribution has its own checksums and third-party notices. Historical source-review reports describe the original 91-file snapshot and are not altered to imply that a different source was reviewed.

The dynamic and static manifests declare different executable SHA-256 prefixes, `d7235178` and `9d05f40d`. They share declared configuration prefix `7e3ecb06`. Source/build hashes are unset, so the difference cannot be certified as launcher-only and the inspected reference source cannot be bound to either executable. The source review identifies CPU/GPU semantic differences in Sobel, Histogram and partial-Gaussian paths. These are properties of inspected code, with historical deployment unresolved. No per-frame equivalence test was supplied. Even for Median’s compatible algorithmic intent, device-level numerical equivalence is unmeasured.

Table S13: Companion files and their roles. These files are supplied in Online Resource 2; basename groups are shown below. The preserved analysis retains its `comparison/`, `erl_analysis/` and `static_analysis/` subdirectories.

Artifact group	Purpose
<code>ESM_1.tex</code> , <code>tables/</code> , <code>figures/</code> <code>build_supplement.py</code>	Self-contained PDF source, table fragments and vector plots. Rebuilds fragments from the parent analysis and preserved evaluation; does not modify primary inputs.
<code>authoritative_18run_inventory.csv</code>	Full mission identifiers, declared binary/configuration hashes, record counts and local raw coverage.
<code>full_recording_inventory_validation.json</code>	Package hash checks, raw-subset binding and exact rechecks.
<code>final_results_verification.json</code>	Input hashes, aggregation definitions, totals and summary-recalculation checks.
<code>final_results_run_workload_medians.csv</code>	All 72 primary within-run workload medians.
<code>final_results_cell_centres.csv</code>	All 36 two-run centres with observed min–max ranges.
<code>final_results_mission_statistics.csv</code>	Full-precision per-mission operation, timing and integrated-energy quan- tities.
<code>strict_gap_*.csv</code>	Fixed static bounds, floors, weight vectors and score-admission sensitivi- ties.
<code>phase_block_bootstrap.csv</code>	Preserved within-phase bootstrap outputs, not independent-replicate in- tervals.
<code>final_forecasts_validation.json</code>	Event-projection hashes, exclusions, chronology checks and forecast scope.
<code>final_forecasts_coverage.csv</code> , <code>final_forecasts_baseline_run.csv</code>	Clean target counts and identical-target forecast comparisons.

All claims remain conditional on recorded policies, observed inputs and retained measurements. Temperature constraints differ from threshold-adherence checks; enabled GPU differs from performed useful work; integrity differs from measurement validity; and retained state differs from a proven learning benefit. Neither missing moderate statics nor source-binding/output-equivalence evidence is replaced by an assumption. The descriptive trade-offs and state continuity are reproducible from the supplied analysis within these boundaries.