

Precision Bluetooth Timing on ESP32-S3: Causal Clock Discipline and Hardware Event Capture

Jonathan Yantis

Independent researcher

22 October 2026, measured-results follow-up

Research draft; not submitted or peer reviewed

Abstract

An implemented ESP32-S3 receive-capture backend completed a fixed 660 s downstream relative-PPS experiment with 657 eligible observations, three invalid records, zero nominal missing opportunities and an observed post-window fence. The actual eligible mean was -106.811 ns, uncentered RMS 166.590 ns, standard deviation 127.842 ns and maximum absolute residual 1591 ns. Four retained excursions above $1\text{ }\mu\text{s}$ account for 39.302% of the uncentered squared error. The result meets the finite collection-coverage objective but fails the sub-100 ns output target. A separate prespecified sparse comparison scored 62 of 64 held-out opportunities: empirical fine-marker RMS was 171.004 ns versus 259.324 ns for matched legacy descriptors. Near-50 ns internal causal clock-quality readings remain distinct from delivered downstream PPS error. The backend joins exact retained packets to RX-active falling edges on one continuing 80 MHz counter; its RF endpoint is unqualified and TX remains a programmed software estimate. The earlier final-profile short legacy check (48 eligible, 99.608 ns RMS) and incomplete legacy 660 s attempt retain separate populations and missingness. The new 660 s population subsequently selected a provisional $51,016$ ns composite offset and cannot validate it. A fresh same-offset N31 transfer retained 48 eligible and 16 invalid records, with 150.954 ns RMS and $+115.167$ ns mean, failing the output target. This report documents the measurements, fault repairs, reproducible accounting and concrete remaining work on TX capture, RX endpoint/delay, tails and fresh transfer validation.

Copyright © 2026 Jonathan Yantis. Paper, figures and derived evidence: CC BY 4.0. Code and ordinary documentation: MIT.

1 Measured findings and scope

The main completed observation is a heterogeneous BLE path: an upstream Nordic source, an ESP32-S3 N30 receiver/estimated relay and a downstream Nordic T096 witness. The N30 has no GNSS. The witness evaluates its delivered BLE map at a locally captured physical PPS; it does not measure an N30 output pin. Its absolute PPS-to-UTC association is unavailable. Every accuracy-like value here is signed fractional-second, source-relative phase unless explicitly identified as an internal or counter-comparison metric.

Three findings govern interpretation. First, the new capture backend and repaired firmware sustained a complete finite 660 s exported record population; the previous collection ending had failed. Second, packet-associated marker capture improved a separate matched causal comparison, while its uncentered RMS still exceeded the target. Third, the complete delivered PPS result retained bias and four large tails. Neither finer timer units nor the internal fit score explains those tails away.

The publication/relay correction was frozen at $50,909$ ns before the completed window. All eligible phase values, the three invalid records, the window boundary and terminal fence are retained in the selected supplement. No retrospective offset, channel correction, residual cut or tail removal changes the reported 166.590 ns RMS. Reusing this population afterward to choose $51,016$ ns changes its role for the next profile, not its original measured result.

The earlier reviewed edition remains preserved in Git and its original receipt [4]. This revision reports implemented changes and completed observations; it does not pool different source paths, boots, estimators or reference conventions into one benchmark. The following pages begin with output coverage and tails, then show calibration, the implemented capture path, the controlled comparison and the remaining measured limitations.

2 The complete 660-second record

2.1 Coverage was recovered; the output target was not

The prespecified interval uses the witness’s nominal 16 MHz native counter, inclusive start and exclusive end separated by exactly 660 s in one local capture epoch. All 660 in-window records were exported: 657 eligible, three reason-2 invalid, zero sequence gaps and zero duplicates. A post-window sequence fence was observed. Nominal 1 Hz accounting is $660 = 657 + 3 + 0$ eligible, invalid and unrecorded opportunities; eligible coverage is 99.545%. This is an opportunity model, not an independent physical count of all pulses.

Eligible edges span 659.000357 s. Mean is -106.811 ns, uncentered RMS 166.590 ns and standard deviation 127.842 ns. Absolute P50/P95/P99 are 100/239/310 ns, with signed range -1591 to 190 ns. The complete trace retains the eligible extremes and breaks at invalid rows; their stored numeric values are not scored.

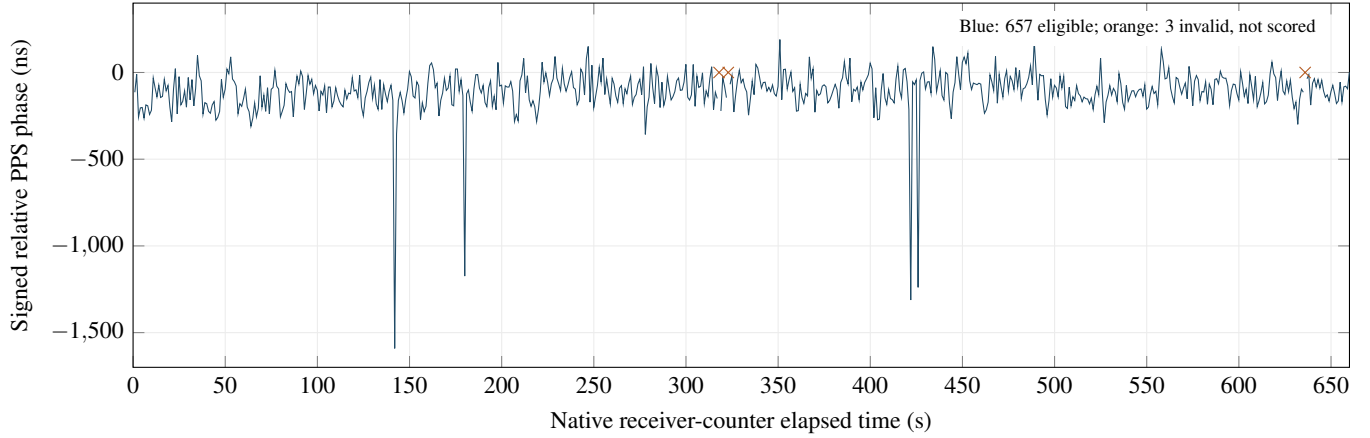


Figure 1: All retained phase values on the full prespecified 660 s axis. No fitted validation offset or detrending is applied. Four eligible excursions exceed $1\ \mu\text{s}$. Invalid crosses show stored phase only; they do not improve the residual score.

Table 1: The four largest eligible excursions, retained in every headline statistic. Relative times use the original fixed-window origin.

Elapsed s	Signed phase ns	Elapsed s	Signed phase ns
141.951515	-1591	179.951535	-1174
421.951664	-1312	425.951667	-1239

These four records are 0.609% of the eligible population and contribute 39.302% of its uncentered squared error. Seventy-three of 657 eligible residuals exceed 200 ns in absolute value. This accounting describes the concentration of error; no trimmed score replaces the full result. The cause of the excursions is not isolated by this experiment.

Three unread terminal edges occurred *after* the observed post-window fence and are not in-window missing records. Queue-drop and disabled-edge deltas were zero. The collection lasted 664.023 s after its native start, including fence/drain. Counter checks, transport accounting and empty polls are described with the method and supplement; a complete software export does not establish calibrated UTC or RF endpoint accuracy.

3 Calibration, short checks and historical coverage

3.1 Fine calibration was frozen before the completed window

The fine-backend calibration run retained 51 records: the first 48 eligible plus three reason-2 invalid. It had no interior gaps and one terminal unread edge. With nominal programmed-to-advertised totals of 39,999 or 40,000 ns, mean phase was $-10,908.833$ ns, RMS 10,909.108 ns and standard deviation 77.382 ns. That centered spread is not actual corrected accuracy. Rounding the training mean selected an additional 10,909 ns and a 50,909 ns total, before the next experiment.

Nine complete selected packet arithmetic pairs, with 18 individual CRC envelopes, verified the installed total before the completed 660 s observation; none was rejected in that proof. The calibration run's separate proof also had nine pairs and 18 valid envelopes, plus one rejected/incomplete envelope. Integer map projection can make the nominal term differ by one nanosecond; the advertised total is the identity check. It is a composite of the path and reference convention, not a calibrated RF delay.

Table 2: Distinct actual populations; RMS includes bias. The first 48 fine observations are a subset of the complete window. No populations are pooled.

Population	Eligible	Mean ns	RMS ns	Std ns	Max $ e $ ns
Legacy 46,993; fresh short check	48	-2.083	99.608	99.586	241
Legacy 46,993; incomplete 660 s	615	42.346	110.801	102.390	402
Fine nominal; calibration only	48	-10908.833	10909.108	77.382	11127
Fine 50,909; first 48 subset	48	-143.542	168.941	89.089	284
Fine 50,909; completed 660 s	657	-106.811	166.590	127.842	1591
N31 51,016; fresh transfer48	48	115.167	150.954	97.590	362

3.2 A good short check does not bound the next full window

The final legacy 46,993 ns profile's fresh 48-sample check had mean -2.083 ns, RMS 99.608 ns and maximum 241 ns. It belongs to the descriptor/GPTimer backend and a different root/path and boot. Its next long attempt retained 615 eligible and one invalid, with 44 nominal unrecorded opportunities after diagnostic collection disabled. The original missing post-window fence and FAILED outcome are preserved. Last record time was 615.679 s, leaving 44.321 s unavailable at the end. Physical PPS failure was not established; conditional errors cannot bound that ending.

The new completed window recovers exported finite coverage, but has higher conditional RMS and larger tails. Its first 48 eligible observations had RMS 168.941 ns, standard deviation 89.089 ns and maximum 284 ns; later tails are absent from that subset. Different backend, source path, boot and window length prevent a controlled platform or offset-effect ranking against the legacy short check. The old lower short RMS remains a historical observation, not the accuracy floor of the new backend.

3.3 Offset selection and a fresh board transfer

After the completed score was frozen, all 657 eligible observations were reused to select a further +107 ns, giving a 51,016 ns total. This makes them calibration evidence for the next profile. They cannot verify it. Even perfect mean removal on the same cohort leaves 127.842 ns RMS, its observed standard deviation, exceeding the 100 ns goal. A constant offset cannot remove its spread and tails. A fresh second-board N31 transfer on the same total yielded 48 eligible of 64 records, with 16 reason-2 invalid, zero sequence gaps/duplicates and one terminal unread edge. Mean was +115.167 ns, RMS 150.954 ns, standard deviation 97.590 ns and maximum 362 ns over a 63.000031 s eligible span. The total was not mean-zero portable; the sub-100 ns target failed. No retuning or N31 TX packet-log proof is part of this separate short observation.

4 Implemented capture path and measurement controls

4.1 Exact packet association precedes timing authority

The N30's experimental backend routes controller receive-active signal RW_RX_ON 238 through GPIO4 into MCPWM group 1. GPIO5 observes TX-active signal 237, but supplies no qualified TX timestamp. The driver reports an 80 MHz capture clock, nominally 12.5 ns per raw count. Espressif documents S3 capture clock/resolution behavior [1]; the queried clock rate does not calibrate RF timing accuracy.

The packet journal retains complete original Report6 advertising data and its CRC, original Stamp/anchor, actual PDU length, extension byte and descriptor status. Selection is deterministic, independent of phase error, RSSI or a favorable width. All candidate edges, fences and late invalidations remain recorded. The production candidate admits only its supported class: 191-octet PDU, raw PHY 0, extension length zero, 190 complete AD octets and one unambiguous RX rise/fall pair. Missing history, multiple candidates, reused edges, unsupported metadata and late callbacks remain unavailable or fault cases.

The descriptor/controller bracket uses a configured $\pm 3100 \mu\text{s}$ search radius plus the retained descriptor enclosure. It does not assign fine receive phase. The selected raw fall is translated by an empirical $1568 \mu\text{s}$, subtracted before rounding to the delivered 50 ns coordinate, nearest with ties upward. This is an implementation reference convention. RF END/ADDRESS and gate-to-radio delay remain unknown; the timestamp is not renamed a calibrated packet endpoint.

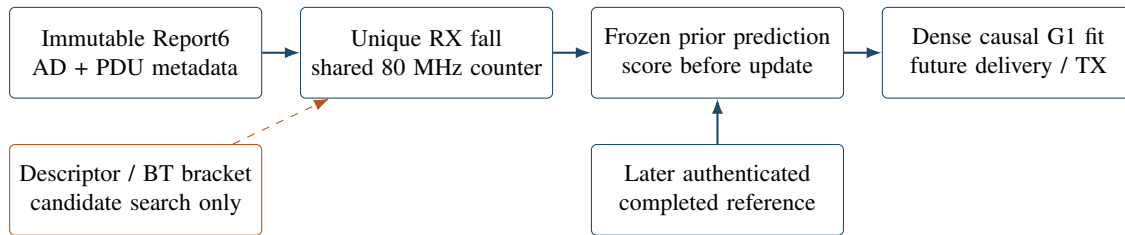


Figure 2: Implemented event/reference flow. The coarse bracket identifies candidates without supplying receive phase. Original capture prediction is retained before later completed truth trains the clock. TX is still estimated.

4.2 One continuing counter; explicit export accounting

One boot-owned extended counter now serves RX, CURRENT, reference availability/completion, Bluetooth-clock mapping and TX estimates. There is no mixed-counter fallback. Nominal 32-bit wrap is 53.6870912 s, half-wrap 26.8435456 s. Stale events, counter restart, bad rate/progress or incarnation mismatch withdraw timing authority. The existing dense 64-point causal discipline and source-expiry behavior remain; a later reference cannot improve the prediction previously frozen at capture.

The witness receives the relay while retaining its existing $+10,543 \text{ ns}$ normalization. Eligibility requires enabled diagnosis, valid map, reference seen/fix/PPS, frozen exact scope and reason zero. Invalid numeric phases remain exported. In the completed window, captured dropped/expired/truncated/refused/write/frame-error deltas were all zero. The audit retained 403 N30 raw-CRC and 1565 witness CRC frames; 763 not-found responses were empty polls. RAM overwrite count rose 64 to 728 as ring retention advanced; with zero cursor gaps this is not missing exported data. A capture latch is not an event FIFO, so clean callback/export sequences alone cannot prove every physical transition survived before interrupt service.

5 Controlled marker comparison and internal quality

5.1 Discovery favored RX fall, without qualifying the endpoint

The first eight deterministic windows retained 16 edges and eight clean windows. Seven had exact common packets and one did not. No late invalidation or audit missing record was observed; maximum audit-query gap was 0.634 s. On the seven pairs, an in-sample affine raw-counter comparison gave 55.523 ns RMS for RX fall, 42.738 μ s for RX rise and 169.588 ns for legacy descriptors. Counter origin and drift were fitted. Those discovery residuals selected a hypothesis; they are not held-out causal output accuracy.

The independent Nordic audit joins the same complete AD using wire peer/root boots and epochs, train, publication, event, channel and AD CRC, with actual path identities checked separately. Device-local capture epochs are not compared as if shared. Correction-completion records and S3-trained projected nanoseconds do not substitute for raw radio capture. CRC proves byte identity, not authentication; the Nordic RX audit producer follows authenticated decoding.

All paired discovery packets used the same PHY/PDU/header class. They cannot establish the fall's RF END/ADDRESS meaning or behavior under other lengths, empty windows and failed packets. The fine-specific composite was calibrated anew rather than inheriting the legacy path correction.

5.2 A prespecified causal comparison improved but missed 100 ns

The next 128 deterministic selections were split into 64 training and 64 held-out opportunities. Across all selections, 127 had unique marker windows, 122 exact common packets and six unpaired windows; one training marker window was ambiguous. The held-out population had 62 scored pairs, two unpaired and zero unrecorded. No replacement or residual-based cut selected a better population.

Table 3: Matched held-out scores from the prespecified sparse last-eight-prior/minimum-four affine comparator. This is distinct from production dense G1 and downstream PPS.

Property	Fine marker	Matched legacy
Held-out scored / unpaired	62 / 2; zero unrecorded	
Causal mean (ns)	−6.807	−7.247
Uncentered RMS (ns)	171.004	259.324
Maximum absolute residual (ns)	452.233	642.966
Paired span / local wraps	332.998 s / 6	
Audit missing / late invalidated	0 / 0	

Conditional RMS improved 34.1% for matched held-out scores, while remaining above 100 ns. Reference-availability ordering was fixed; no later truth entered a prior prediction. Incomplete channel coverage and the separate sparse estimator limit generalization. The result supports a causal marker improvement on this population, not a 171 ns bound for production G1 or actual output.

5.3 Internal precision and a repaired resource fault

The completed N32 research session records three finite internal causal Sync RMS readings of 49 ns among 311 finite readings in 313 snapshots, with steady readings around 53 ns. These are capture-time pre-update prediction-quality values, not downstream PPS error. The 313 normal status snapshots include acquisition and are exported separately; their rolling RMS is neither a whole-run aggregate nor continuous-lock proof.

An earlier low-internal-RMS image later encountered HCI 0x07 Memory Capacity Exceeded during advertising-data update. The tested revision removes unused mutually exclusive WINDOW storage (about 7488 allocated DRAM bytes) and repairs maintenance clock-shutdown ordering. The completed cohort retained one counter incarnation with no reported counter fault, late invalidation or timeout, and sampled healthy status to at least 761.056 s. This is finite post-repair evidence, not long-term reliability certification.

6 Observed bottlenecks and comparison limits

6.1 The short prefix did not contain the later large errors

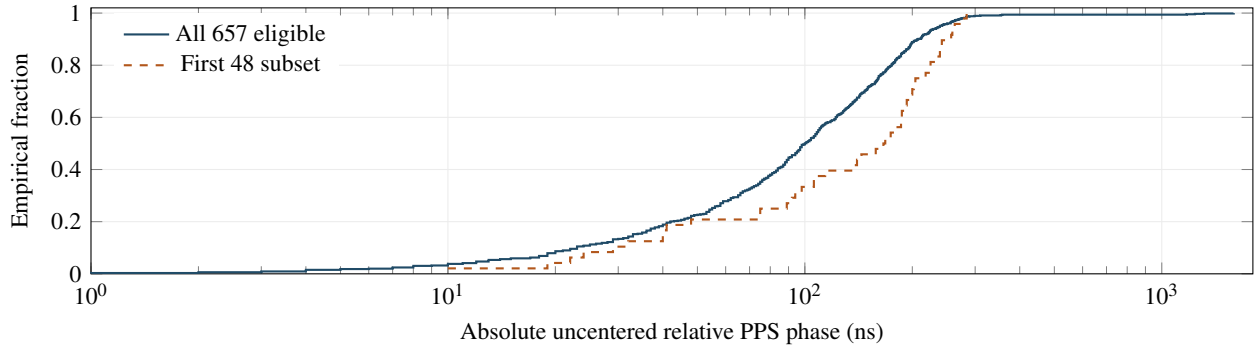


Figure 3: Actual uncentered residual distributions. The full curve reaches 1591 ns; the first 48 are a subset, not an independent trial. Zero is represented at the left logarithmic plotting boundary. Three invalid rows are excluded from residual curves and counted separately.

The short subset’s 89.089 ns centered spread did not bound the full cohort’s 127.842 ns spread. The largest excursions appeared after the first two minutes and again near seven minutes. Neither the earlier 77.382 ns training spread nor a later mean correction qualifies a sub-100 ns result. There is no evidence here that a constant offset removes the mechanism behind the four tails.

6.2 TX remains a software estimate with measured bracket width

The relay advertises a reference derived from a programmed scheduling point, not an actual RF TX latch. The implementation reads four Bluetooth-clock brackets, selects the narrowest, uses its midpoint and projects the programmed point through a frozen map. Retained proof brackets span approximately 7.9–9.25 μ s; the controller field retains 500 ns quantization. One continuing local counter removes a mixed-domain phase bridge, but does not turn that software estimate into physical TX time.

Nominal ADDRESS-convention conversion is 40 μ s; integer map projection can yield 39,999 or 40,000 ns. Proof pairs verify that nominal and residual terms sum to the installed total. They do not bound scheduling, detector, RF or propagation error. Bracket midpoint latency/asymmetry and quantization are plausible contributors to boot bias or tails, not a proven sole cause. Counter origins from different boots are not comparable phase measurements.

6.3 Nordic history retains its own denominators

Nordic nRF52840 exposes RADIO ADDRESS, TIMER capture and PPI event routing; its nominal 16 MHz grid is 62.5 ns [2]. That defined hardware event is an endpoint advantage over the S3’s empirical receive-active fall and estimated TX. Detector and antenna delay still require calibration; nominal tick size is not output accuracy.

The preserved V14 one-relay population has 602 nominal opportunities: 436 reference-valid causal, 65 invalid, three expired and 98 unobserved. Conditional mean is -18.417 ns, RMS 62.504 ns, standard deviation 59.729 ns and maximum 191 ns, at 72.425% nominal eligible coverage. Its 501-sample all-causal set including invalid references has 3.929 ms RMS and 19.753 ms maximum; the largest witness gap is 59.017 s [3]. The new S3 population has greater finite eligible coverage and larger conditional error. Different source paths, dates, estimators and validity patterns preclude a controlled SoC ranking. The older Nordic result is context, not a target already attained by the S3.

7 Concrete remaining improvements

The measured priorities are to replace estimated TX, establish the RX reference convention's physical delay, investigate tails and test transfer on a fresh population. The existing common counter is retained throughout. The completed full-window statistics, rather than a favorable prefix or training spread, define the work still required. No proposed mechanism is counted as a result.

7.1 Verify actual TX and its association

Freeze the exact selected advertising bank and programmed-map snapshot independently of bulk logging. Correlate the actual transmitted packet with a local event latch and an independent Nordic raw ADDRESS record, retaining no-match, multiple-match and late cases. GPIO5 TX-active is only a candidate signal. A qualified TX timestamp requires both exact packet identity and an independently established RF endpoint/delay. The existing nine-pair proof establishes advertised arithmetic alone.

A new latch must live in the same counter domain as RX, CURRENT, completion and availability. It must preserve source expiry, counter-fault withdrawal and actual reference-availability ordering. Demonstrating narrower software brackets without measuring the event would leave the current physical TX uncertainty unresolved.

7.2 Resolve RX delay and investigate the tails

Repeat RX rise/fall correlation with independent raw events while varying real PHY/PDU/header lengths, and include empty, failed and unrelated-traffic windows. The present supported class alone cannot separate RF END, gate delay and timeout behavior. An affine intercept absorbs counter origin and fixed path delay; a small fitted residual therefore cannot calibrate phase. Retain the empirical convention and UNKNOWN endpoint where alternatives remain indistinguishable.

Investigate the four retained large excursions using immutable packet, map-activation and counter-health records. Test candidate causes prospectively, rather than removing the affected samples or renaming them invalid after observing their error. A fixed offset may reduce bias, but the observed mean-only floor is already above 100 ns. Any change intended to reduce variance or tails requires new held-out output observations with fixed scoring and coverage rules.

Table 4: Next measurable decisions. These are uncompleted validation tasks, not predicted improvements.

Change or check	Required observation before claiming success
Actual TX capture	Same transmitted AD, native local latch and independent RF-correlated endpoint; all unmatched/ambiguous cases retained.
RX endpoint/delay	Varied PHY/length and failed/empty controls; delay variation and independent timing uncertainty stated, not absorbed into an unexplained constant.
Clock-map delay	Event capture, map activation and completion availability share the continuing counter; no later reference trains an earlier score.
Tail reduction	Fresh fixed-population output comparison including every eligible extreme, invalid record and missing opportunity; no evaluation-set offset fit or residual cut.
Sub-100 ns / 62.5 ns	Treat these as distinct delivered-relative-PPS RMS targets with maximum and coverage reported. A timer grid or internal fit does not pass either target.
Broader transfer	Additional boots/boards on a frozen profile, with identity/scope and advertised arithmetic established before scoring. The completed N31 short check retained a nonzero mean; keep unavailable attempts explicit.

The final 51,016 ns N31 transfer retained a +115.167 ns mean and 150.954 ns RMS. Its 64 records include 16 invalid and no software sequence gaps. No further experiment or second long run is included beyond this final short transfer. A quick transfer check can expose a large board/boot bias; it cannot establish long-term tails or calibrated physical output. Subsequent improvements must be judged on a new population after the mechanism changes.

8 Reproduction and practical limits

8.1 What the reported error measures

The witness projects its delivered BLE map at a locally captured PPS and wraps the signed fractional phase into $[-0.5, 0.5)$ s. PPS-to-UTC association is zero and absolute-error output unavailable. Shared upstream errors can cancel; fixed receiver, detector, normalization and propagation biases may remain. No N30 electrical PPS output or calibrated absolute UTC error is measured. One-second sampling misses subsecond excursions.

For eligible phases we report signed mean, uncentered RMS, population standard deviation, nearest-rank absolute quantiles and absolute maximum. Invalid records retain their flags and numeric values but do not enter those statistics. No demeaning, detrending, residual cut or independent-sample confidence interval is used. Sequence order, duplicate conflicts and the post-window fence audit native-window export; callback completeness cannot establish RF-latch completeness.

One finite source/path and supported packet class cannot qualify other boots, boards, RF environments, temperatures, distances, lengths or PHYs. Sampled status and a finite healthy session do not prove uninterrupted lock or long-term service. The sparse comparison, dense internal quality, downstream PPS and historical Nordic population have different estimators and denominators and remain separate.

8.2 Public reconstruction and edition identity

The selected supplement [5] retains all 51 fine-calibration, 660 fixed-window and 64 N31-transfer records, including 22 invalid rows across those populations. It also retains 313 sampled normal snapshots and provenance-bound association/comparison summaries. Private hardware/root identities are reduced to scope-validity flags; enrollment, raw logs, firmware images, absolute epochs, USB identifiers and private paths are excluded. Hashes establish byte identity, not accuracy.

Run `python3 tools/verify_s3_fine_paper.py` from a public checkout. Standard-library Python recomputes eligibility, residuals, tails, quantiles, minute blocks, the first-48 subset, offset arithmetic and denominators, then checks selected source/PDF/manifest hashes. No hardware, credentials or sibling private checkout is required. Discovery/sparse summaries are selected aggregates; the public export cannot reconstruct their original RF extraction. The previous reviewed edition and manifest remain unchanged in Git; its original verifier runs at the recorded historical revision. This draft's receipt records the matching source/PDF, compiler and all-page render review.

The hard result is a complete finite window with 166.590 ns RMS and a 1591 ns maximum, alongside a separate matched causal marker gain. The measured output targets remain unmet. The implemented common counter and exact association are useful foundations; verified TX, resolved RX delay, prospective tail reduction and fresh transfer evidence remain the next concrete work.

References

- [1] Espressif, *ESP-IDF v5.5.5 Programming Guide: MCPWM, ESP32-S3*, capture timer resolution and event capture. <https://docs.espressif.com/projects/esp-idf/en/v5.5.5/esp32s3/api-reference/peripherals/mcpwm.html>.
- [2] Nordic Semiconductor, *nRF52840 Product Specification*, RADIO ADDRESS, PPI and TIMER capture. https://docs.nordicsemi.com/r/bundle/ps_nrf52840/page/radio.html; https://docs.nordicsemi.com/r/bundle/ps_nrf52840/page/timer.html.
- [3] J. Yantis, *Passive Precision Time over BLE: V14 Boundary Clocks, Causal Timestamping, and Prospective One-Relay Validation*, research draft, 12 October 2026. Historical evidence: <https://github.com/yantisj2/PToB/tree/main/evidence/v14/2026-10-12>.
- [4] J. Yantis, *Precision Bluetooth Timing on ESP32-S3*, first reviewed draft, 22 October 2026. Immutable manifest `papers/editions/s3-timing-2026-10-22.json`; original evidence `evidence/s3/2026-10-22`. Source/PDF preserved at the revision recorded in the follow-up receipt.
- [5] J. Yantis, *ESP32-S3 empirical fine-capture follow-up*, selected derived evidence, 22 October 2026. `evidence/s3/2026-10-22-fine; papers/editions/s3-timing-2026-10-22-fine.json`. Original paper, figures and datasets CC BY 4.0; code and ordinary documentation MIT. Research draft, not submitted or promoted.