

Certification Resources and the Value of Retained Quantum-Reader Records

Jeremy Rodgers
Independent Researcher
www.everythingequation.com

27 October 2026

<https://doi.org/10.5281/zenodo.23721122>

Abstract

We characterize task-relative validation resources for frozen quantum readers. Terminal-error observations admit intersection–union and sequential likelihood gates with nearly matching expected-stock information bounds. In a common trusted shared-visibility family, readable trajectories strictly improve those resources: a frozen fourteen-copy reader can be validated with false-certificate probability 2^{-20} and nominal acceptance probability at least .95 using at most 720160 source copies, whereas every terminal-error-only validator under the same physical-family promise needs more than 1992371 in expectation. At the exact nominal device, an uncapped trajectory gate accepts almost surely with expected stock below 254746; a quantified conditional response tube retains a similar expected separation. We also prove an unrestricted transcript obstruction and construct an exact trusted-seed information-projection test. The family, preparation and deployment-stability promises remain assumptions; terminal-risk validation does not establish full instrument equivalence.

1 What is being certified

This paper develops task-relative validation and retained-record resource comparisons within Shadow Theory’s pre-existing source–readout programme [6]. The results concern the declared reader and validation interfaces. A reader is frozen before validation. Each complete B -copy experiment returns a sign at its deadline, retaining all settings, signs, nulls, root seeds and declared timings. A trusted evaluator knows the source label for scoring; the reader does not. The deployment contract is the nominated maximum label error ϵ , not recovery of a hidden source or characterization of all unread outputs. For the structured comparisons, the informative trajectory alphabet is the root seed, commanded actions and ternary outcomes $(+, -, 0)$. Declared timing fields are determined by these records or have the same source- and model-independent conditional law in all compared worlds. They add no likelihood factor. A source- or visibility-dependent arrival-time observation would enlarge the statistical experiment and requires its own response model; it is not supplied by the declared effects or by a different detector construction. We separate three interfaces: unrestricted terminal Bernoulli risks; richer records with specified trusted ancillary laws; and a shared physical response model whose trial likelihood can be evaluated. Their information minima need not agree. Both sides of each resource comparison receive the same preparation and response promises. Validation stock, policy/controller state, archive access, clocks, phase references and arithmetic are separate resources. Unless a finite-cap theorem explicitly says otherwise, a finite expected stopping cost does not imply a finite worst-case duration. The concrete reader is the 697-node frozen rational fourteen-copy DAG in the supplement, with action/three-child pointers and terminal signs. Its two fresh uniform root bits implement sign reversal and axis exchange. The source is fixed within an experiment among $\rho_{X\pm} = (I \pm X)/2$, $\rho_{Y\pm} = (I \pm Y)/2$; the target is its common sign. Copies are independent conditional on that label, not

an independently redrawn axis mixture. The physical effects are $E_{\pm|\theta} = \eta(I \pm v[\cos \theta X + \sin \theta Y])/2$, $E_0 = (1 - \eta)I$. General sections below do not assume that family; the structured comparison explicitly assumes known $\eta = .8$ and fixed unknown shared v .

1.1 Conventions, conditioning and information access

Readable validation records in this paper are classical records of a quantum reader; access to unread residual quantum systems is not supplied. Unless specialized below, $0 < \alpha < 1$, $0 < \epsilon < 1$, the experiment horizon B and label count H are positive integers, and a planning risk satisfies $0 < p_0 < \epsilon$. A power parameter obeys $0 \leq \beta < 1 - \alpha$. We use natural logarithms and

$$\text{kl}(p||q) = p \log(p/q) + (1 - p) \log[(1 - p)/(1 - q)], \quad \text{TV}(P, Q) = \sup_A |P(A) - Q(A)|,$$

with the usual limiting convention for zero event masses. For discrete laws, $\text{TV}(P, Q) = \frac{1}{2} \sum_z |P(z) - Q(z)|$. One validation observation is a complete B -copy experiment; source-copy cost is B times the number of such observations.

The proof filtration contains completed experiment records and their mathematically defined scoring truths. During trusted validation the evaluator can use the known preparation label; the reader remains adapted only to its permitted record subfiltration. Conditional source, response and tube promises are imposed on this augmented past, including admitted timing metadata. They do not follow merely by enlarging a smaller observed-history filtration. In the structured four-state comparisons, each new experiment draws an independent uniform root seed conditional on the completed past and its chosen label. The unrestricted-record and general trusted-seed models state their own ancillary-law promises instead. Structured block drift averages over its fresh seed; local likelihood validity is pointwise once the seed and command are known. No future validation-success event is conditioned on as initial information.

Organization. Section 2 gives terminal-risk gates, and Section 3 their restricted information floor. Section 4 separates unrestricted records from trusted-seed access. Section 5 proves the structured trajectory advantage and finite cap; Section 6 supplies a separate conditional response-tube guarantee. The appendix compares validation resources for two specified deployed readers.

1.2 A complete frozen-policy interface

The finite DAG is a mathematical input, not a controller that can inspect the hidden label. At a node i with t trials left, its action is one rational unit direction $a_i = (c_i, s_i)$ and its three outcome edges lead to nodes $j_{i,o}$ with $t - 1$ trials left. At $t = 0$, the node stores a sign d_i . With $R(h)$ the nominated sign, define

$$r_{0,i,h}(v) = \mathbf{1}\{d_i \neq R(h)\}, \quad r_{t,i,h}(v) = \sum_{o \in \{+, -, 0\}} K_v(o \mid h, a_i) r_{t-1,j_{i,o},h}(v).$$

Here $K_v(\pm 1 \mid h, a) = \eta(1 \pm v x_h)/2$, $K_v(0 \mid h, a) = 1 - \eta$ and $(x_{X+}, x_{X-}, x_{Y+}, x_{Y-}) = (c, -c, s, -s)$. The recursion is a conditional calculation, not label access by the controller. It evaluates the exact conditional error including every null. The root's independent sign/axis symmetries give $q(v) = \frac{1}{4} \sum_h r_{14,\text{root},h}(v)$ for each physical label under the shared-response model. All actions and edges are supplied in `supplement/readers/research12/rich/policy_14.json`; the byte-identical reader input also appears in `research11/nominal/` within that supplement. The input manifest fixes its identity. Consequently the risk and likelihood proofs below need neither an optimality assertion about that policy nor a theorem promised by another manuscript.

Safe likelihood testing, reverse information projection and sequential change-of-measure have established antecedents[2–4, 7]. Our contribution is their precise operational integration, matched comparison and exact certificates within the declared source–readout programme. No worldwide priority, empirical validation or Born-selection claim is made.

2 Task-relative validation with a near-optimal expected stock bound

Response-level calibration can certify more than one frozen reader needs. We validate its four terminal-error probabilities directly. Exact binomial gates give a finite-budget interface; likelihood-ratio gates give an anytime interface with a sharp expected-stock comparison. These are established testing methods, newly applied and fully proved for the declared source–reader contract, not a new quantum probability law.

2.1 The validation interface and indispensable stability promise

Fix a complete reader policy Π with horizon B , root randomness, null handling and final decoder before opening validation data. A trusted calibration controller requests one of the four source labels h , supplies B fresh copies independent conditional on that fixed label, and records the terminal error $Z = \mathbf{1}\{\Pi \text{ decides the wrong sign}\}$. The label is available to the external auditor for scoring, but is not an input to the reader policy. Every ordered setting, sign, null and controller seed remains retained; no record is postselected.

The physical stability contract is that the repeated complete experiments, conditional on the fixed device and trusted label, have one error probability q_h and fresh independent stock. Predictably choosing the next label from past calibration data is allowed, provided the conditional error law remains q_h . Deployment must have the same conditional policy-error probabilities. This promise is not inferred from reset, past calibration, or a detector’s average density equation. Drift, changed preparation, unread environment memory, policy changes and reference depletion require separate allowances.

The target is $q_{\max} = \max_h q_h \leq \epsilon$. The gate reads only the terminal error bits and trusted label identifiers for its decisions. Richer archives are appended but not queried. Thus it certifies one frozen policy’s loss, not a continuous response curve, arbitrary future interventions or a complete retained quantum instrument. This restricted information interface is also the class of the matching stock lower bound in Section 3.

2.2 Finite exact gates without a four-label Bonferroni penalty

For n complete experiments on label h , let S_h be its error count and write $F_{n,p}(c) = \Pr\{\text{Bin}(n, p) \leq c\}$.

Theorem 2.1 (Intersection–union policy certificate). *Choose fixed label budgets n_h and count thresholds c_h with $F_{n_h, \epsilon}(c_h) \leq \alpha$. Accept the frozen policy only if $S_h \leq c_h$ for every label. Uniformly over all fixed risk vectors with $q_{\max} > \epsilon$, the probability of issuing this false certificate is at most α . The per-label test size is α , not $\alpha/4$.*

Proof. Pick any label h_* whose true risk exceeds ϵ . The all-pass event is a subset of $\{S_{h_*} \leq c_{h_*}\}$. Binomial lower-tail probability is nonincreasing in its success parameter, by coupling Bernoulli variables with common uniforms. Therefore its probability is at most $F_{n_{h_*}, \epsilon}(c_{h_*}) \leq \alpha$. No union over the labels is needed: the null is a union of bad-label hypotheses and acceptance is an intersection of their individual rejection events. Independence between label-gate events is unnecessary for this soundness argument. $\square$

For a nominated n , the largest integer c satisfying this inequality is the most permissive deterministic lower-count gate. It maximizes acceptance power within that count-threshold class for every $q < \epsilon$. The usual Bernoulli monotone likelihood ratio gives the broader Neyman–Pearson interpretation, with optional randomization at a boundary if that larger class is admitted. We do not claim that the selected finite budget below is the globally minimum finite budget over all sequential designs. Exact binomial inversion is established statistics[1, 5]; the stated inequality is independently implemented here.

At $B = 14$, $\epsilon = 1/100$, $\alpha = 2^{-20}$, take

$$n_h = 365000, \quad c_h = 3366 \quad (h = X+, X-, Y+, Y-).$$

Integer binomial sums certify

$$F_{365000,.01}(3366) < 2^{-20} < F_{365000,.01}(3367).$$

Thus the chosen count threshold is maximal at this budget. If every true $q_h \leq p_0 = 887/100000$, the same exact calculation gives $F_{365000,p_0}(3366) > 79/80$. A union bound on failure then gives all-label acceptance probability greater than $19/20$, without using an independence assumption between the four gate events. This consumes

$$4 \cdot 365000 = 1460000 \text{ whole experiments,} \quad 14 \cdot 1460000 = 20440000 \text{ source copies.}$$

The planning risk $p_0 = .00887$ is a rational upper bound on the nominal fourteen-copy bisector error $.008866347\dots$; its exact quadratic-field comparison is replayed. This is a synthetic good-device planning case, not an observed validation run.

The minimum sample count allowing deterministic acceptance of a zero-error record under the same per-label size is exactly 1380, because $(.99)^{1380} \leq 2^{-20} < (.99)^{1379}$. That is a minimum for accepting that record, not a useful high-power design at p_0 . Short calibration that can accept a very rare good-looking record is different from a procedure likely to certify a genuinely good reader.

2.3 Anytime validation with certified expected cost

Choose an alternative $0 < p_0 < \epsilon$ and define for each label

$$E_h(n) = \left(\frac{p_0}{\epsilon}\right)^{S_h(n)} \left(\frac{1-p_0}{1-\epsilon}\right)^{n-S_h(n)}.$$

Query labels predictably. Mark a label accepted irreversibly the first time $E_h \geq 1/\alpha$; stop querying it. Issue the policy certificate only when all labels are marked. Testing one unaccepted label at a time is one simple implementable allocation. Nulls within each B -copy experiment still enter its terminal error and remain archived.

Theorem 2.2 (Anytime intersection–union gate). *The probability of ever issuing a false policy certificate is at most α under every fixed risk vector with $q_{\max} > \epsilon$, including predictable label allocation and stopping. For the attaining allocation, process one unaccepted label at a time until it crosses, then continue through the remaining labels, without early rejection. If every $q_h \leq p_0$, this serial exhaustive rule accepts all labels almost surely, and for each label’s crossing query count τ_h ,*

$$\mathbb{E}\tau_h \leq \frac{A+b}{D}, \quad A = \log(1/\alpha), \quad b = \log \frac{1-p_0}{1-\epsilon}, \quad D = \text{kl}(p_0\|\epsilon) > 0.$$

Consequently that complete procedure has expected source consumption at most $4B(A+b)/D$. Arbitrary predictable allocation, early rejection or stopping remains permitted for false-certificate control; it does not by itself carry this acceptance/attainment promise. No finite deterministic completion time is asserted.

Proof. For a bad label with $q_h \geq \epsilon$, the expected multiplier on its next query is

$$q_h \frac{p_0}{\epsilon} + (1 - q_h) \frac{1 - p_0}{1 - \epsilon} = 1 + \frac{(q_h - \epsilon)(p_0 - \epsilon)}{\epsilon(1 - \epsilon)} \leq 1.$$

Its nonnegative likelihood-ratio process, held constant when another label is queried, is therefore a supermartingale in the global calibration filtration. Stopping it at its threshold and applying the elementary nonnegative-supermartingale maximal inequality gives crossing probability at most α . Any false all-label certificate requires a crossing by one fixed bad label, proving the intersection–union claim. The underlying anytime testing machinery is established[3]; it is proved here for this particular gate and filtration.

For the expected-cost claim, use the serial exhaustive allocation and put $a = \log(p_0/\epsilon) < 0$. On each good label query the log multiplier has expectation $q_h a + (1 - q_h)b \geq D$, since $a < b$ and $q_h \leq p_0$. The log process stopped at $\tau_h \wedge N$ is bounded above by $A + b$: before crossing it is below A , and an upward crossing adds only b . Taking expectations of the finite stopped increment sum gives $D \mathbb{E}(\tau_h \wedge N) \leq A + b$. Monotone convergence proves the displayed finite expected bound. A positive probability of infinite τ_h would make that expectation infinite, so each label crosses almost surely. Summing four label expectations proves the source-stock bound. $\square$

At $p_0 = .00887$, $\epsilon = .01$, $\alpha = 2^{-20}$ and $B = 14$, outward rational logarithm enclosures give

$$\mathbb{E}\tau_h < 206802.173725, \quad \mathbb{E}[\text{source copies}] < 11580921.729.$$

For the frozen adaptive fourteen-copy policy, its separate risk certificate permits the smaller planning bound $p_0 = .00875$. With the same target, confidence and horizon,

$$\mathbb{E}\tau_h < 168277.131290, \quad \mathbb{E}[\text{source copies}] < 9423519.353.$$

The matched expected upper budget decreases by more than 18.6%. A finite-command implementation whose separate phase-error certificate gives risk at most .0088 instead has expected source stock below 10243551.605. That third entry is conditional on its independently checked physical-command allowance, rather than silently treating real-valued commands as free. The smaller planning risk is not assumed for an arbitrary uncalibrated device; it defines the good-device class on which the acceptance-cost promise holds. False-certificate control remains valid outside that class. The second planning case is used only after the frozen policy’s mathematical risk bound is independently checked.

Section 3 proves the matching terminal-score lower bound $\mathbb{E}N_h \geq A/D$ for every validator with false-certificate size α and almost-sure acceptance at the all- p_0 good vector. Our per-label upper is within the factor

$$1 + \frac{b}{A} < 1.0001$$

of that lower frontier in both planning cases. This is a restricted expected observation-cost statement, not a universal optimum over physically structured detector models, informative quantum archives, controllers or calibration preparations. A trusted physical relation between different label risks may permit a different frontier.

3 An information lower boundary for terminal-risk certification

A direct decision-risk gate has a weaker output contract than identifying an entire continuous detector response. We can nevertheless bound its best possible resource cost in a precise black-box interface, and compare an attainable sequential gate with that boundary. This distinguishes genuine task-relative economy from a claim of universal physical-calibration savings.

3.1 The declared validation experiment

A frozen reader uses B fresh source copies and returns a sign at its fixed deadline. A trusted evaluator prepares one of H labels, keeps its identity unavailable to the reader, and scores the terminal error $Y \in \{0, 1\}$. The calibration algorithm may select labels predictably, randomize, stop or reject, but its allocation and acceptance use only these label/error observations. Reader settings, nulls and root seeds remain append-only and unread by the validator; its own allocation coins are independent of the reader seed law. Querying additional trajectory information changes this lower-bound model.

Conditional on a fixed device and label h , successive calibration experiments have independent Bernoulli error probability p_h . This is an explicit fresh-stock, preparation and stability promise, not a conclusion from reset. The target is $p_h \leq \epsilon$ for every label. An acceptance event A must satisfy

$$\sup_{p: \max_h p_h > \epsilon} \Pr(A) \leq \alpha.$$

For a good boundary device $p_h = p_0 < \epsilon$ for every label, require $\Pr_{p_0}(A) \geq 1 - \beta$, with $0 \leq \beta < 1 - \alpha$. Let N_h be the total number of validation experiments queried at label h , possibly random and sequential. Each such experiment consumes B source preparations; trusted-label setting, controller, references, resets, archives and time are additional resources.

Theorem 3.1 (Sequential terminal-validation lower boundary). *Every validator in the declared interface satisfies*

$$\mathbb{E}_{p_0} N_h \geq \frac{\text{kl}(1 - \beta \| \alpha)}{\text{kl}(p_0 \| \epsilon)} \quad \text{for each } h,$$

and consequently expected source consumption is at least BH times the right-hand side. Here kl is binary relative entropy in natural logarithms. For power one ($\beta = 0$), the numerator is $\log(1/\alpha)$. The bound allows adaptive label allocation and acceptance-only procedures that need not terminate under a bad device.

Proof. Fix a label h . Compare the good vector to an alternative changing only that label's error to $q > \epsilon$. At a finite global query budget T , the likelihood chain rule gives transcript relative entropy

$$D(P_{p_0}^T \| P_{q,h}^T) = \mathbb{E}_{p_0} N_h(T) \text{kl}(p_0 \| q),$$

where stopped algorithms are extended by a no-query action. Adaptive label choices and ready coins have the same conditional controller law in the two worlds and add no direct relative entropy. The event $A_T = \{\text{accepted by query } T\}$ is measurable in this finite transcript. Data processing gives the corresponding binary event-relative-entropy bound.

The sequential change-of-measure principle is established in Kaufmann et al. [4], Lemma 1 and Appendix A.1; the finite-prefix form here also covers a bad world in which acceptance never occurs. Since $\Pr_{q,h}(A_T) \leq \alpha$ and $\Pr_{p_0}(A_T) \uparrow \Pr_{p_0}(A) \geq 1 - \beta > \alpha$, monotonicity and continuity of binary relative entropy imply, as $T \rightarrow \infty$,

$$\mathbb{E}_{p_0} N_h \text{kl}(p_0 \| q) \geq \text{kl}(1 - \beta \| \alpha).$$

Infinite expected consumption makes the inequality automatic. Otherwise monotone convergence justifies the finite limit. Let $q \downarrow \epsilon$ to respect the strict bad-device condition. Sum the resulting bounds over all labels and charge B preparations per experiment. No Bonferroni factor is introduced into the lower bound; every one-bad-label alternative must separately be controlled by the same global acceptance rule. $\square$

This is a lower bound for an unrestricted terminal-Bernoulli validation model. Its one-bad-label alternatives need not all be realizable within a specified low-dimensional Pauli detector family. Therefore it does not lower-bound every structured detector-calibration method, a protocol using full quantum records, or a method with extra known preparation information. It also does not assert that all physical apparatus resources equal sample count. Those boundaries are necessary to interpret the comparison fairly.

3.2 A nearly attained expected-resource boundary

Section 2 constructs the likelihood process for each label,

$$E_{h,n} = (p_0/\epsilon)^{S_{h,n}} [(1-p_0)/(1-\epsilon)]^{n-S_{h,n}},$$

and, under its serial exhaustive allocation, accepts only after every label has crossed $1/\alpha$ at some time. Its intersection–union guarantee uses α per label without dividing by H , while controlling the global false-certificate event at α . Under every good device with $p_h \leq p_0$, the same frozen rule accepts almost surely. Put

$$A = \log(1/\alpha), \quad b = \log \frac{1-p_0}{1-\epsilon}, \quad D = \text{kl}(p_0 \parallel \epsilon).$$

Only a non-error observation increases the log likelihood and its increment is b . Thus its stopped log value never exceeds $A + b$. Positive drift at least D , applied first at a truncated stopping time, proves $\mathbb{E}N_h \leq (A + b)/D$ and then almost-sure acceptance. Combined with Theorem 3.1 at $\beta = 0$, the matched frontier is

$$\boxed{BH \frac{A}{D} \leq C_{\text{opt}} \leq BH \frac{A+b}{D}}$$

for expected source consumption at the good boundary, in this terminal-only interface. The attained upper rule has a stronger uniform good-device promise $p_h \leq p_0$; the lower uses its included boundary vector. The multiplicative gap is at most $1 + b/A$. There is no finite worst-case stopping budget in this expectation theorem; a finite-budget alternative is supplied separately in Section 2.

For $H = 4$, $B = 14$, $\epsilon = .01$ and $\alpha = 2^{-20}$, the declared nominal planning cases are:

Good-risk promise	Expected source lower	Attainable expected upper
$p_0 = .00887$ (bisector planning)	11579968.82...	11580921.73...
$p_0 = .00875$ (adaptive planning)	9422661.68...	9423519.35...

The exact upper figures and outward enclosures are supplied by the calibration verifier. These are mathematical planning bounds, not measured validation costs. The nominal design certificates justify the two chosen risk margins under their calibrated model; a real deployment still needs stable trusted validation. Different p_0 values are explicitly different good-device planning promises. No uniform superiority across all noisy apparatus families follows merely from this table.

The smaller error of the frozen fourteen-copy DAG therefore has a concrete possible benefit despite an unchanged copy upper threshold: it increases the margin to one percent and lowers the cost of terminal-risk certification. Controller/feedback time and actual reference/calibration setup must still be charged. Appendix A supplies the corresponding comparison conditions rather than inferring time from preparations alone.

3.3 Exact logarithmic enclosure and verification

The supplied `research11/calibration/calibration_gate.py` and `research11/calibration/verify_calibration.py`, relative to `supplement/readers/`, evaluate outward rational enclosures of all

displayed information quantities. To enclose $\log x$, reduce x by powers of two to $u \in [1, 2)$ and use

$$\log u = 2 \sum_{j=0}^{m-1} \frac{z^{2j+1}}{2j+1} + R_m, \quad z = \frac{u-1}{u+1}, \quad 0 \leq R_m \leq \frac{2z^{2m+1}}{(2m+1)(1-z^2)}.$$

The same series encloses $\log 2$; signed integer multiples are rounded in the proper direction. The exported rational cost intervals are then rounded outward to twelve decimal places for readability. Floating displays never choose a certified inequality. Independent rational range reduction and high-precision checks supplement the analytic proof.

4 What richer records can and cannot certify

A full trajectory is a richer observation than its terminal error, but the number of recorded fields alone does not imply a cheaper certificate. This section supplies the missing interface-dependent lower-bound comparison. It also gives an exact trusted-seed repair. These are specializations of classical change-of-measure and safe-testing methods, not new general statistical principles; the programme contribution is a checked account of which source–readout promises permit a resource reduction.

4.1 An unrestricted full-record no-free-lunch result

Let Z be the entire classical record of one completed frozen B -copy experiment, including any available commands, seed and outcomes, and let $Y = f(Z) \in \{0, 1\}$ indicate a wrong terminal sign. At label h its good law is P_h , with $p_h = P_h(Y = 1) \in (0, \epsilon)$. Suppose the admitted model of record laws is closed under changing the error mass while keeping both conditional record laws $P_h(\cdot | Y)$ fixed. This is an explicit unrestricted observation model, not a claim about the Pauli apparatus.

Theorem 4.1 (Error-mass tilt preserves the terminal information floor). *For $q \in (p_h, 1)$ define*

$$\frac{dQ_{h,q}}{dP_h}(z) = \begin{cases} q/p_h, & f(z) = 1, \\ (1-q)/(1-p_h), & f(z) = 0. \end{cases}$$

Then $Q_{h,q}$ is a probability law with error q and

$$D(P_h \| Q_{h,q}) = \text{kl}(p_h \| q).$$

Consequently every sequential validator with access to the full Z , global false-acceptance probability at most α whenever some label error is above ϵ , and good-world acceptance probability at least $1 - \beta > \alpha$, obeys

$$\mathbb{E}_P N_h \geq \frac{\text{kl}(1 - \beta \| \alpha)}{\text{kl}(p_h \| \epsilon)}.$$

Here completed experiments are fresh and independent conditional on their selected labels; label allocation and stopping may be adaptive. The bound allows an infinite stopping time in a bad world.

Proof. The two density values normalize because the two event masses under P_h are p_h and $1 - p_h$. Integrating their logarithms gives the binary relative entropy exactly. Change only label h to this alternative, with $q > \epsilon$. At a finite global query cap the transcript KL chain rule gives $\mathbb{E} N_h(T) \text{kl}(p_h \| q)$; allocation coins have the same conditional law in both worlds. Data processing through acceptance by that cap, followed by monotone convergence as the cap increases and then $q \downarrow \epsilon$, proves the result. This is the same finite-prefix argument used in the earlier terminal-only analysis and the established sequential change-of-measure literature. $\square$

Thus merely reading an unconstrained archive does not defeat the previous information limit. A useful improvement must exploit a restriction on the alternative record laws, not their representation length. Conversely, closure under this tilt must not be silently assumed for a physically specified detector. The tilt can violate a trusted seed marginal, a known erasure law or a declared response family. Such restrictions are genuine additional information and can yield a different frontier.

4.2 A trusted-seed information projection and attaining gate

Assume each experiment begins with an independently trusted seed $S \in \{1, \dots, m\}$, with known fixed weights $w_s > 0$, $\sum_s w_s = 1$. The validator observes S and Y after the complete experiment. It neither selects a seed retrospectively nor discards an inconvenient seed. Deployment uses the same seed law. The seed weights hold conditional on the completed validation past and selected label; completed experiments use the declared fresh stock. Thus the following likelihood and drift statements are conditional on the common proof filtration, without giving the reader access to the hidden preparation label. For one label, nominate a good conditional-error vector $p_s \in (0, 1)$ with $\bar{p} = \sum_s w_s p_s < \epsilon < 1$. The bad class consists of all conditional risks q_s whose average exceeds ϵ , with no hidden response-model constraint.

Theorem 4.2 (Seed-informed lower boundary and safe likelihood gate). *There is a unique $q^* \in (0, 1)^m$ minimizing*

$$I_w(p, \epsilon) = \min_{\sum_s w_s q_s \geq \epsilon} \sum_s w_s \text{kl}(p_s \| q_s).$$

It satisfies $\sum_s w_s q_s^* = \epsilon$ and, for one $\lambda > 0$,

$$q_s^* - p_s = \lambda q_s^* (1 - q_s^*), \quad q_s^* = \frac{\lambda - 1 + \sqrt{(1 - \lambda)^2 + 4\lambda p_s}}{2\lambda}.$$

For the weights trusted above, define

$$f_s(1) = p_s / q_s^*, \quad f_s(0) = (1 - p_s) / (1 - q_s^*).$$

The product of these observed multipliers is a nonnegative supermartingale under every bad conditional-risk vector. Stop and certify the label when it first reaches $1/\alpha$. Under the exact good vector p , or under a coordinatewise smaller conditional-risk vector, it crosses almost surely with expected experiment count bounded by

$$\mathbb{E} N_h \leq \frac{\log(1/\alpha) + b_w}{I_w(p, \epsilon)}, \quad b_w = \max_s \log f_s(0).$$

Every seed-informed validator with false-certificate probability at most α and good acceptance at least $1 - \beta$ instead satisfies

$$\mathbb{E}_p N_h \geq \frac{\text{kl}(1 - \beta \| \alpha)}{I_w(p, \epsilon)}.$$

For multiple labels, an intersection–union gate may certify only when every label passes; no confidence split over labels is required.

Proof. The objective is strictly convex in q for interior p . Its unconstrained minimum is p , which is infeasible, and moving any $q_s < p_s$ upward toward p_s improves the objective and the constraint. The optimum is therefore at the average-risk boundary with $q_s > p_s$. Differentiation gives the common positive multiplier equation. The displayed positive quadratic root increases continuously from p_s to 1 as λ increases, so there is one multiplier giving the nominated boundary.

The same equation implies

$$f_s(1) - f_s(0) = -\lambda, \quad \sum_s w_s f_s(0) = 1 + \lambda\epsilon.$$

Hence at any alternative vector q ,

$$\mathbb{E}_q f_S(Y) = 1 + \lambda \left(\epsilon - \sum_s w_s q_s \right) \leq 1.$$

This proves composite-null validity, including predictable label allocation. Ville’s inequality gives the crossing-size bound. At p , expected log increment is I_w , and reducing any p_s increases that drift since $f_s(1) < f_s(0)$. Before crossing the log product is below $A = \log(1/\alpha)$; an upward step adds at most b_w . Applying the finite stopped increment sum yields $I_w \mathbb{E}(N_h \wedge n) \leq A + b_w$, and monotone convergence proves finite expectation and almost-sure crossing.

For the lower bound, preserve the seed weights, choose errors just above q^* , and preserve all other record shapes conditional on (S, Y) . Per-experiment divergence decreases to I_w as that alternative approaches the boundary. The finite-prefix acceptance change-of-measure proof above then applies. An all-label certificate with a bad label requires that particular bad label’s crossing; this proves the intersection–union claim. $\square$

This construction specializes established reverse-information-projection safe testing[2] and sequential change-of-measure bounds[4] to the declared reader-seed interface. The contribution is the explicit specialization and verification, rather than a new general information projection or likelihood-testing principle.

Log-sum/data processing gives

$$I_w(p, \epsilon) \geq \text{kl}(\bar{p} \parallel \epsilon).$$

For interior, heterogeneous p_s the inequality is strict. Equality would require both p_s/q_s^* and $(1 - p_s)/(1 - q_s^*)$ to be constant across seeds. Two distinct p_s then force both constants to one, contradicting $\bar{p} < \epsilon$. Known seed marginal information can therefore improve the terminal-only information floor. It does not ensure a large improvement for a particular reader.

4.3 Actual fourteen-copy check: seed heterogeneity is too small

The verified unsymmetrized fourteen-copy base risks, in label order $X+, X-, Y+, Y-$, are approximately

$$(.008642797684269258, .009021802859698374, .008473934585934470, .008858815447399504).$$

For each fixed physical label the four trusted root symmetries permute this vector. Thus every label has the same mean $.008749337644325401 \dots$. The code constructs a nearby exact rational good-design vector dominating the independently verified outward bounds, and a rational boundary vector with mean exactly $.01$. It chooses one rational λ so the quadratic relation is exact. This avoids presenting numerical root finding as a proof.

The resulting certified information ratio is

$$\frac{I_w}{\text{kl}(\bar{p} \parallel .01)} = 1.00000546458 \dots$$

It implies only about 0.000546% improvement in the matched information floor. The upper expected source-stock account is below 9.413292 million for four labels and fourteen copies per complete experiment. This is not the earlier $.00875$ rounded planning case: comparisons must use the same

good-risk promise. Root-seed stratification alone is therefore not a meaningful practical advance for this frozen reader.

An exact two-seed illustration uses weights $(1/2, 1/2)$, $q^* = (.001, .019)$ and $\lambda = 1250/9819$. It has mean good error .00875 and an information ratio 1.00789355.... This checks the strict-gain construction but is not claimed as a new quantum-reader protocol.

Reading an even richer archive can help under a justified detector-response model, as in the separate structured protocol. It is that restricted model, not a universal claim that more fields always lower certification cost, that must support any substantive savings. All tests here certify terminal risk only; neither the unread bath nor the complete retained quantum instrument is identified. Expected stopping costs have no finite worst-case completion bound. A selected policy must be frozen before fresh validation, or selection multiplicity must be controlled separately.

5 Readable trajectories can reduce certification stock under a declared response model

The terminal-score information floor is not a universal calibration floor. We construct a full-record test and compare it with *every terminal-score-only test within the same physical family*, rather than charging the latter for four unrelated label risks. This supplies a substantial attainable resource separation while making the additional model promise explicit.

5.1 One common physical and preparation contract

Freeze the shipped fourteen-copy DAG and its fresh two-bit uniform root symmetrization. The device has known $\eta = 4/5$ and one fixed unknown visibility $v \in [0, 1]$. Every commanded rational direction (c, s) has the promised response

$$K_v(o \mid h, c, s) = \frac{\eta}{2}(1 + ovx_h), \quad o = \pm 1, \quad K_v(0 \mid h, c, s) = 1 - \eta,$$

where $x_h \in \{c, -c, s, -s\}$, with the recorded root transformation applied. This shared sinusoidal law, trusted commands, independent fresh source copies conditional on one fixed hidden label, independent fresh detector stock and stability between calibration and deployment are explicit premises. A reset does not prove them. The evaluator can prepare one nominated trusted label (for example $X+$), hidden from the reader, and knows it for scoring and likelihood evaluation. Calibration consumes the actual complete fourteen-trial experiments, including all nulls.

Under the response promise, fresh root symmetry equalizes all four label risks to one function $q(v)$, the mean of the four base-DAG risks. This does not assert that four risks are equal for an arbitrary device. Terminal-only comparators receive the same physical-family promise and may choose labels adaptively; their observations are terminal errors only and their allocation cannot inspect reader seeds or intermediate records. The richer evaluator additionally reads those records after each complete experiment. The informative record here is the seed, commands and ternary outcomes; timing fields obey the common noninformative timing contract stated in the introduction. Both retain the same deployment law and may randomize their validation algorithms. No outcome feedback is added to a reader counted as nonadaptive.

5.2 A uniformly certified good-visibility region

For the frozen DAG, exact outward integer evaluation on

$$v_j = .89 + j/10000, \quad 0 \leq j \leq 1100,$$

finds the largest grid risk upper bound below 0.009612708713. This is not taken as evidence for global monotonicity. For any two visibilities, a local outcome-kernel TV distance is

$$\frac{\eta}{2}|x_h||v-w| \leq \frac{\eta}{2}|v-w|.$$

History-wise coupling or telescoping across the fourteen adaptive trials therefore bounds the terminal-risk difference by $14\eta|v-w|/2$. Every point of $[\cdot 89, 1]$ lies within $1/20000$ of a grid point. Adding the resulting allowance $7/25000 = .00028$ gives the exact certified upper bound

$$\sup_{.89 \leq v \leq 1} q(v) \leq \frac{145990612645647230665109}{14757395258967641292800000} < .009892708713 < .01.$$

Thus every bad device in this family, meaning $q(v) > .01$, has $v < .89$. This sufficient good region need not coincide with the true risk boundary. The grid cover includes every omitted visibility and every adaptive history; it makes no assumption about the reliability of a floating-point optimizer.

5.3 A stronger whole-interval risk certificate

An independent Bernstein calculation proves more than the grid cover for this frozen reader. For $v = u + (w - u)z$, write each remaining- t risk polynomial in Bernstein form of degree t . A child has degree $t - 1$, and every local kernel is affine in z . If $b_{i,h,k}$ denotes the node's coefficient, its exact recursion is

$$b_{i,h,k} = \frac{t-k}{t} \sum_o K_u(o | h, a_i) b_{j_o,h,k} + \frac{k}{t} \sum_o K_w(o | h, a_i) b_{j_o,h,k-1},$$

with out-of-range terms omitted. Constant terminal errors initialize the degree-zero coefficients. The formula follows from multiplication of degree- $t - 1$ Bernstein basis polynomials by $1 - z$ and z ; all weights are nonnegative. Outward dyadic rounding therefore gives rigorous coefficient intervals without polynomial cancellation. The root's four-label mean is again taken outward.

For $[u, w] = [0, 1]$, all fourteen adjacent root coefficient pairs are strictly decreasing even after the rounding intervals are charged. The derivative identity

$$q'(v) = \frac{14}{w-u} \sum_{k=0}^{13} (b_{k+1} - b_k) \binom{13}{k} z^k (1-z)^{13-k}$$

then proves that this particular policy's risk is strictly decreasing throughout $[0, 1]$. It does not prove monotonicity for arbitrary frozen policies, robust policies or optimized protocol classes. The separate $[\cdot 89, 1]$ certificate has maximum coefficient below 0.009612708713, strengthening the grid upper. The gate and misspecification figures below retain the weaker grid allowance conservatively, so they do not depend on using this sharpening.

5.4 An anytime full-record likelihood certificate

Put $v_0 = .9$ and $v_c = .89$. After each complete experiment multiply the running statistic by its trajectory likelihood ratio

$$Z = \prod_{t=1}^{14} \frac{K_{v_0}(o_t | h, a_t)}{K_{v_c}(o_t | h, a_t)}.$$

The actual commands are predictable functions of earlier outcomes and the recorded seed. Their conditional controller probabilities cancel between the two worlds. Seed probabilities likewise cancel. A null contributes one. Declare acceptance after the first completed experiment at which the cumulative likelihood reaches $1/\alpha$, with $\alpha = 2^{-20}$.

Theorem 5.1 (Structured full-record gate). *Under the stated shared-response family, the probability of ever accepting a device with $q(v) > .01$ is at most α . At $v = .9$ the gate accepts almost surely, with certified expected source consumption below 254746 copies. Uniformly over $v \geq .9$ it accepts almost surely with expected source consumption below 587860 copies. Neither bound is a finite worst-case stopping budget.*

Proof. For a signed projection $x = x_h$, direct rational algebra gives the one-step likelihood multiplier expectation at actual visibility v :

$$\mathbb{E}_v \left[\frac{K_{v_0}(o \mid h, a)}{K_{v_c}(o \mid h, a)} \middle| \mathcal{F} \right] = 1 + \frac{\eta(v - v_c)(v_0 - v_c)x^2}{1 - v_c^2 x^2}.$$

It is at most one for every $v \leq v_c$. Therefore the product is a nonnegative supermartingale for all bad devices in the family. The nonnegative-supermartingale maximal inequality controls an anytime crossing by α , even with queries restricted to full-experiment boundaries.

At v_0 , each trial's expected log multiplier is its conditional kernel divergence

$$d(a, h) = \eta \operatorname{kl} \left(\frac{1 + v_0 |x_h|}{2} \parallel \frac{1 + v_c |x_h|}{2} \right).$$

Bottom-up DAG expectation with the v_0 transition probabilities, followed by the four virtual-label average induced by the root seed, encloses the mean experiment log increment as

$$D_{\text{path}} = 0.0007659246736557862 \dots$$

The implementation encloses every logarithm by rational series bounds. This figure includes all source-copy consumptions, nulls and outcome-dependent commands; it is not fourteen times one independently optimized local information value.

For every $v \geq v_0$, the local expected log increment is at least $d(a, h)$, since increasing v increases the expectation of the log likelihood ratio. To avoid using nominal path frequencies in an unknown world, recursively define $G_0 = 0$ and

$$G_t(i, h) = d(a_i, h) + \min_o G_{t-1}(j_{i,o}, h).$$

The root's uniform virtual-label average supplies the conservative uniform experiment drift lower bound $D_{\min} > 0$, explicitly recorded in the exact output. This recursion remains valid for the changing distribution of histories as v varies.

Every trial's positive log increment is at most $\log(190/189)$; a whole experiment adds at most $b = 14 \log(190/189)$. Let $A = \log(1/\alpha)$. The stopped cumulative log is bounded above by $A + b$. At each finite truncated stopping time, summing conditional log drifts gives $D\mathbb{E}(N \wedge M) \leq A + b$. Monotone convergence then gives finite expectation and almost-sure acceptance. Substitution of D_{path} or $D_{\min}$ and multiplication by fourteen proves the stated budgets. $\square$

The risk cover leaves approximately .00010729 below the target, but no unmodelled phase or calibration error is silently charged against it: such an extension would also need to preserve or robustify the likelihood supermartingale. The present theorem assumes the displayed response law and trusted delivered commands. It does not certify arbitrary command-dependent errors or an unread residual quantum instrument.

5.5 A same-family lower bound for terminal-score validators

Take the good visibility $v = .9$ and the admitted bad alternative $v_b = 177/200 = .885$. Exact DAG intervals give

$$q(.9) = .008749337644325401\dots, \quad q(.885) = .01006704807047499\dots > .01.$$

Every label has the same seed-averaged risk in each world. Thus adaptive label selection, stopping and validator randomness contribute no extra information to a terminal-only transcript. The relative entropy per queried experiment is precisely $\text{kl}(q(.9)||q(.885))$.

If a terminal-only validator has false-acceptance probability at most α throughout this same family and acceptance probability at least .95 at $v = .9$, the stopped chain rule and binary data processing imply

$$\mathbb{E}_g[\text{source copies}] \geq 14 \frac{\text{kl}(.95||\alpha)}{\text{kl}(q(.9)||q(.885))}.$$

For almost-sure acceptance the numerator becomes $\log(1/\alpha)$. The finite-prefix proof of the earlier terminal-only change-of-measure bound applies directly; it does not require termination in the bad world. The reported exact output encloses these expressions using monotonicity in the two narrowly enclosed risks.

The directed calculations give a power-.95 terminal-only lower bound

$$\mathbb{E}_g[\text{copies}] > 1992371.13869,$$

whereas the readable-record gate has upper bound 254745.042770. Its expected stock is therefore more than 87.21% below the lower bound for every admitted terminal-only validator. The power-one terminal-only lower bound is 2129329.265776...

Consequently the readable-trajectory gate beats an information lower bound for every terminal-only comparator satisfying the same physical-family promise and the same or weaker good-world power requirement. This is an information-access benefit inside a declared physical family, not merely comparison against an inefficient terminal-score test. This comparison does not transfer to the unrestricted Bernoulli validation interface of Section 3: its larger floor uses different alternatives and cannot serve as the matched comparator here.

5.6 Finite calibration deadline with certified power

Expected stopping cost alone need not supply an economical fixed calibration deadline. Here a separate exact affinity calculation gives such a deadline under the exact shared-response parent at nominal $v = .9$.

Let P_0 and P_c be the complete fourteen-trial transcript laws at $v_0 = .9$ and $v_c = .89$, with the root seed, settings and all outcomes retained. Define their affinity

$$M = \sum_{\omega} \sqrt{P_0(\omega)P_c(\omega)}.$$

There is no exponential transcript enumeration. At a remaining- t DAG node, initialize terminal affinity to one and recur as

$$H_t(i, h) = \sum_o \sqrt{K_{v_0}(o | h, a_i)K_{v_c}(o | h, a_i)}H_{t-1}(j_{i,o}, h).$$

Because the root seed is retained and has the same uniform law in both worlds, the full affinity is exactly the four-virtual-label mean of the root affinities. Replacing this by the affinity after discarding the seed would generally give a different, larger value.

Dyadic square-root intervals propagate outward through this nonnegative recursion and prove

$$M = 0.9998070301369535 \dots < 1.$$

After N independent complete calibration experiments at the nominal world, with L_N their un-stopped full likelihood ratio,

$$\mathbb{E}_0 L_N^{-1/2} = M^N.$$

If the likelihood gate has not crossed $1/\alpha$ by experiment N , then $L_N < 1/\alpha$. Markov's inequality consequently gives

$$\Pr_0\{\text{not accepted by } N\} \leq \alpha^{-1/2} M^N.$$

Directed logarithm enclosures certify that $N = 51440$ satisfies the right-hand side at most $1/20$ for $\alpha = 2^{-20}$. Stop validation at that deadline, issuing no certificate if it has not crossed. This capped rule therefore has false-certificate size at most 2^{-20} throughout the declared family, nominal acceptance power at least .95, and deterministic source-copy cap

$$\boxed{14 \cdot 51440 = 720160.}$$

This is more than 63.85% below the 1992371 *expected* copy lower bound for every terminal-only validator with the same physical-family promise and nominal power requirement. The comparison is stricter than comparing two expected upper bounds: the rich rule's finite cap is below a universal comparator's mean lower bound. It is not a universal optimum over richer validation methods.

Independent fresh whole-experiment stock is used for the product identity. The finite-cap power result is nominal and is not silently transferred to the response-error tube below. Deadline rejection is a validation decision, not abstention in the fixed-horizon reader task. The deployment reader still returns a sign after every fourteen trials.

5.7 A controlled response-error tube

The exact response premise can be weakened in a quantified way. Permit the actual outcome kernel at every reached history to lie within TV distance $\zeta = 10^{-6}$ of K_v , for the same fixed unknown v . The allowance holds conditionally on all previous calibration and deployment records, not merely after averaging. It must hold for each trusted label and command; it is not inferred from a few calibration settings.

For an action and label, put $f_o = K_{.9}(o)/K_{.89}(o)$ and replace the factor by

$$\tilde{f}_o = \frac{f_o}{1 + \zeta \text{osc}(f)}, \quad \text{osc}(f) = \max_o f_o - \min_o f_o.$$

The expectation of any bounded function changes by at most its oscillation times TV. Thus the corrected factor remains a supermartingale multiplier for $v \leq .89$. The good-visibility risk region remains safe because

$$q_{\text{actual}} \leq .009892708713 + 14\zeta < .009906708713 < .01.$$

At $v = .9$, let d_i be the nominal local KL lower bound, p_o its nominal kernel, and recursively define rational bounds

$$W_i = d_i - \zeta \text{osc}(\log f) - \log[1 + \zeta \text{osc}(f)] + \sum_o p_o W_{j_o} - \zeta \text{osc}(W_{j_o}), \quad W_0 = 0,$$

with logarithms rounded in the conservative direction. Induction shows this lower-bounds the conditional expected remaining corrected log likelihood against every allowed history-dependent perturbation. The

root's four-seed mean is positive. With the same whole-experiment overshoot bound, the expected source stock is below

$$255043.918928 < 255044.$$

This bound allows predictable kernel perturbations; no fresh independence of those perturbations is assumed. The original exact worlds are contained in the enlarged tube, so the same 1992371 terminal-only lower bound still applies to any validator required to control the tube family. At the exact nominal good device, which is included in the tube, this provides a matched stock separation even when validity is required over the enlarged model. The same terminal-only lower bound is not asserted separately at every perturbed good device. No optimality over tube models is asserted.

5.8 Deployment, resources and limitations

The evaluator needs one trusted source label, a fresh uniform two-bit reader seed per whole experiment, access to all commanded actions and outcomes, and exact or conservative log-likelihood accumulation. Its own acceptance controller is outside the fourteen-copy deployment reader. Counting fourteen source preparations per observation already includes nulls and every complete experiment; history extraction, arithmetic, references, source readiness, reset and clock resources remain additional. Finite worst-case counters and elapsed time are not bounded by this expectation theorem. The reported cost bounds concern the defined exact likelihood crossing. Here its rational kernel ratios permit an exact product and threshold comparison. A conservative approximate log accumulator can preserve false-certificate control, but retaining the same acceptance and cost bounds requires charging its rounding or drift loss separately.

The policy is frozen before calibration. Selecting it using the same records would require a selection-valid uniform gate or fresh validation. If acceptance has probability $P(A)$, the usual joint deployment-error statement is $P(A \cap \text{error}) \leq .01P(A) + \alpha$ under the stability contract. An unqualified conditional-on-acceptance .01 guarantee does not follow without an acceptance-probability allowance. The gate verifies a task-relative risk implication within the trusted physical family, not the family itself, a Born-selection mechanism or a unique source ontology.

6 A predictable response-tube guarantee

The exact sinusoidal family is a physical promise, not something inferred from the observed command archive. A conservative enlargement permits a specified discrepancy without pretending to certify arbitrary response curves. Let each actual conditional outcome law, given the augmented proof past defined in the introduction and the fixed label, be within total variation ζ of the declared kernel K_v at that command. A single fixed visibility $v \in [0, 1]$ applies across calibration and deployment. Ready symmetry bits remain fresh independent uniform conditional on the completed past and chosen label. Blockwise drift is averaged before their new draw; one-trial validity holds after the seed and command are known. This conditional tube is required even across experiment boundaries; neither reset nor past successful validation proves it. It can allow correlated conditional discrepancies, but not violations of the tube itself.

Theorem 6.1 (A safe robustified full-record likelihood). *Write $v_0 = .9$, $v_c = .89$, $\eta = .8$, $B = 14$. For a recorded local outcome use $f = K_{v_0}(o)/K_{v_c}(o)$. Then*

$$10/11 \leq f \leq 190/189, \quad \kappa = \frac{200\zeta}{2079}.$$

Replace each multiplier by $f/(1+\kappa)$, retain all records, and check for crossing $1/\alpha$ at complete-experiment boundaries. Under every $v \leq v_c$ and every allowed conditional tube discrepancy, the resulting product is

a nonnegative supermartingale. Its crossing probability is at most α . If the model risk is uniformly bounded by q_c for $v \geq v_c$, every actual conditional deployment risk there is at most $q_c + B\zeta$.

Proof. For a trusted label, put $x = (c, -c, s, -s)_h$. Nulls have multiplier one. Direct algebra gives

$$\mathbb{E}_{K_v} f = 1 + \frac{\eta(v - v_c)(v_0 - v_c)x^2}{1 - v_c^2 x^2} \leq 1 \quad (v \leq v_c).$$

The two extreme nonnull ratios occur at $x = \pm 1$, giving the displayed range and its oscillation 200/2079. For bounded f , a TV perturbation of size ζ changes its expectation by at most $\zeta \text{osc } f$. Normalization proves the conditional supermartingale inequality; Ville's inequality controls every crossing, including the selected boundary times. Sequential coupling of the B actual outcomes with their model outcomes, using the same controller and ready coins, fails with probability at most $B\zeta$. On the complementary event the terminal sign agrees. The same coupling applies conditional on the past, proving the deployment claim. $\square$

The action-dependent normalizer in Section 5 gives the sharper nominal tube bound below 255044. The present construction uses one global normalizer, independent of the action, and a conservative uniform-good drift bound. It is a different, simpler robustified test; its larger budgets below do not supersede that sharper nominal result.

For expected costs let $A = \log(1/\alpha)$, $b = 14 \log(190/189)$, and $R = \log(209/189)$. Let D_0 be the certified nominal mean log growth per complete experiment, and let $D_{\min}$ be the certified path-minimum lower growth for every $v \geq .9$ in the exact family, as computed in the structured certificate. For an exact nominal good device, the normalized experiment drift is at least

$$d_0 = D_0 - 14 \log(1 + \kappa).$$

For any good device in the entire conditional tube with $v \geq .9$, it is at least

$$d_{\text{tube}} = D_{\min} - 196\zeta R - 14 \log(1 + \kappa).$$

Indeed the complete-experiment log likelihood has range at most $14R$ and the two conditional experiment laws differ by TV at most 14ζ . The normalization is deterministic. If either drift lower bound d is positive, truncating the boundary stopping time gives

$$\mathbb{E}[\text{source copies}] \leq 14(A + b)/d.$$

There is no finite worst-case stopping duration. Positive drift also proves almost-sure acceptance in the stated good class. This argument permits a correlated tube because it lower-bounds each experiment's drift conditional on its entire preceding history.

At $\zeta = 10^{-6}$ and $\alpha = 2^{-20}$, the exact directed-log replay gives

$$q_c + 14\zeta < .009906709 < .01, \quad C_0 < 255194, \quad C_{\text{tube}} < 627692.$$

These are expected source-stock upper bounds; trusted efficiency, preparation, controller, reference and timing contracts remain additional. The conditional TV tube is separately supplied, not established by the likelihood test. The same physical family with zero discrepancy is included in this tube, so the explicit bad visibility used for the terminal-only lower bound remains an admitted alternative. The structured same-family comparison therefore continues to demonstrate a record-access benefit at the exact nominal good device, even when false-certificate validity is required over this enlarged model. This does not extend the claim to arbitrary command-dependent errors.

7 Deployment conditioning, resources and reproducibility

If the same validation data select a policy, fresh validation or a selection-uniform certificate is required. For acceptance event A and stable deployment the valid joint statement is $\Pr(A \cap \text{error}) \leq \epsilon \Pr(A) + \alpha$. Conditional-on-acceptance risk therefore retains an $\alpha/\Pr(A)$ allowance; it is not simply $\epsilon + \alpha$. Neither successful calibration nor reset proves unrestricted future stability. The strict readable-record comparison uses the same deployed fourteen-copy policy. Its effect comes from information discarded by the terminal score, rather than from an adaptive/nonadaptive deployment comparison or four unrelated risk alternatives. Additional archive extraction, exact logarithmic arithmetic, preparation trust, resets, phase references and controller time must be charged. The conditional TV tube is an additional allowed model, not a conclusion learned from a few probes. The finite-cap power theorem applies to independent exact nominal complete experiments, not automatically to correlated tube perturbations. The top-level `reproduce.py` replays the reviewed standard-library checks in disposable copies; `supplement/readers/INPUT_MANIFEST.json` identifies the policy, exact certificates, declared derived input and verifier modules. The scoped job list also includes the appendix’s resource and independent information-cost checks. The structured likelihood, monotonicity and finite-cap checks are in `research12/rich/`, with independent checks in `research12/audit/` and `research12/root/` relative to that supplement. Independent forward occupancy and interval-log calculations check the trajectory drift; ordered-path affinity checks verify the finite cap. The supporting research record includes useful unsuccessful routes, but only checked statements appear as results. AI assistance was used for review and computational checking.

A An explicit validation-resource and timing comparison

This comparison uses the frozen fourteen-copy adaptive policy and the fourteen-copy repeated bisector. It does not optimize over all nonadaptive readers. Both protocols already attain the nominated one-percent deployment target. The extra error margin can nevertheless reduce the cost of verifying that target from trusted whole-experiment trials.

A.1 A checked fixed-reader error enclosure

At the nominal shared detector, a clicked bisector observation has correct-sign probability

$$p = \frac{1 + v/\sqrt{2}}{2}.$$

Conditional on n clicks, the fair-tie majority rule has error

$$f_n(p) = \sum_{j < n/2} \binom{n}{j} p^j (1-p)^{n-j} + \frac{1_{2|n}}{2} \binom{n}{n/2} p^{n/2} (1-p)^{n/2}.$$

Thus its actual error at $B = 14$, $\eta = 4/5$, $v = 9/10$ is

$$q_N = \sum_{n=0}^{14} \binom{14}{n} \eta^n (1-\eta)^{14-n} f_n(p).$$

The success event with a fair terminal tie coin is increasing in all Bernoulli observations. Coupling those observations with common uniform random variables proves that f_n is nonincreasing in p . The verifier brackets $\sqrt{2}$ by adjacent 100-bit dyadic rationals, computes the corresponding endpoint sums exactly, and certifies

$$.008866 \leq q_N \leq .00887, \quad q_N = .00886634707138013 \dots$$

All four labels have this same error for the fixed bisector. This check is independent of the point-based search. The supplied `research11/sources/resource_comparison.py` and `research11/audit/check_resource_comparison.py` reproduce the appendix’s stock, finite-phase and timing inequalities. Their declared `research11/nominal/results.json` input is the already certified frozen policy-risk output, preserved with its exact hash; the reader’s policy and shared-response risk are also independently evaluated by the trajectory checks in this paper.

A.2 A strict comparison within the terminal-risk interface

Set $\epsilon = .01$ and false-certificate probability $\alpha = 2^{-20}$. The adaptive policy’s certified error is at most .00875 for every label after its admitted symmetry randomization. The sequential intersection–union gate gives an expected source-stock upper bound

$$U_A \leq 9,423,519.353$$

under that good-risk promise. There is no finite worst-case stopping budget. The comparator may choose its trusted labels adaptively, stop adaptively, and use a different terminal validation test. Require it to be false-certificate safe over the same unrestricted Bernoulli error alternatives and to accept the actual nominal bisector with probability at least .95. The validation interface exposes only the trusted label and terminal error bit, together with the validator’s own randomness. The experiment’s settings, nulls, controller seed and timing records remain in an append-only archive but are not readable for validation stratification or feedback. The lower bound does not cover a validator admitted to those additional records without a further complete-law change-of-measure argument.

For a one-bad-label alternative whose error tends down to ϵ , the validation information bound gives, for each label h ,

$$\mathbb{E}_N N_h \geq \frac{\text{kl}(.95, \alpha)}{\text{kl}(q_N, \epsilon)}.$$

Since $\text{kl}(q, \epsilon)$ decreases for $q < \epsilon$, the certified lower bound $q_N \geq .008866$ can replace q_N in the denominator conservatively. Summing the four labels and charging fourteen source copies per terminal experiment yields

$$L_N \geq 10,757,305.473.$$

The adaptive upper bound is strictly below this fixed-reader validation lower bound, by at least 12.3988%. If the comparator must instead accept almost surely, its lower bound becomes 11,496,776.338 and the guaranteed fraction is at least 18.0333%.

This is a comparison of expected validation resources at the specified good devices, under a strong black-box validity requirement. It does not prove a universal gain over structured Pauli calibration, other nonadaptive readers, or every physical implementation. Its two good devices have different errors; that difference is the mechanism of the certification saving. The positivity and stopping arguments, not a comparison of two loose cost upper bounds, establish the strict result.

A.3 A finite-precision variant

Suppose actual conditional measurement angles differ from the commanded angles by at most $\delta = 10^{-5}$ radians, with the specified detector parameters and independent preparation contract otherwise unchanged. For any label and accessible history, the one-copy outcome total variation changes by at most $\eta v \delta / 2$. Sequential coupling and the union bound therefore increase any fixed policy’s error by at most

$$\frac{B \eta v \delta}{2} = .0000504.$$

This bound permits conditional angle errors and hence does not require them to be freshly independent. The certified nominal adaptive error plus this allowance remains below .0088. A nearest 19-bit command on the quarter circle has rounding error at most $\pi/[4(2^{19} - 1)] < 1.499 \times 10^{-6}$ radians. The remaining error budget is a separately required actuator and reference promise; the bit count does not demonstrate that promise in hardware. The quarter-circle encoding also requires its coarse quadrant or equivalent outcome/setting relabelling bits. For the subsequent validation-cost comparison, the declared independent whole-experiment terminal-trial contract remains required; a within-experiment perturbation bound alone does not supply fresh independence across repeated experiments.

Using .0088 as the sequential gate’s good-risk parameter gives

$$U_A^{\text{phase}} \leq 10,243,551.605 < L_N.$$

It retains a strict expected terminal-validation stock comparison of at least 4.7758%, against the ideal nominal fixed-bisector validation lower bound with .95 power. A device/model-specific calibration could consume fewer resources than either black-box validation interface, so this is not asserted as a universal physical frontier.

A.4 Timing, setup and reuse

Let a full source trial take τ seconds, including common preparation, detection and reset costs. Charge the adaptive reader up to thirteen additional feedback waits of ℓ seconds per validation experiment. Let s_A be an upper bound on additional per-experiment adaptive phase-switching, controller or reference time; let s_N be an unavoidable lower bound for the comparator. In particular, an upper bound on its switching time cannot be inserted as a compulsory tax. Let D_A be an adaptive setup upper bound and D_N an unavoidable comparator setup lower bound (or their actual fixed setup times). A sufficient condition for a strict expected validation-time improvement is

$$D_A + \frac{U_A}{14}(14\tau + 13\ell + s_A) < D_N + \frac{L_N}{14}(14\tau + s_N).$$

With $D_A = D_N$ and $s_A = s_N = 0$, the certified bounds imply

$$\ell/\tau < .15242554$$

for the ideal-command adaptive gate, or

$$\ell/\tau < .05401187$$

for the finite-phase variant. These are sufficient model-dependent conditions, not measured latency claims. If feedback waits are fewer than thirteen on some paths, the bound is conservative.

After validation, both displayed protocols consume fourteen source copies per deployed decision. Hence this comparison gives a finite validation saving, not a deployment copy saving that grows with the number of uses. If R later deployments consume an extra adaptive overhead $d > 0$ seconds each, the validation-time margin can absorb at most Rd ; any required lifetime or drift guarantee over those R uses remains a separate premise. Conversely, a one-time calibration overhead can be amortized only under such a reuse promise. Append-only archives, causal feedback memory, trusted label preparations, independent ready coins, command precision and reference access remain separately charged resources. No energy or hardware-size inference follows.

References

- [1] C. J. Clopper and E. S. Pearson. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4):404–413, 1934. doi: 10.1093/biomet/26.4.404. URL <https://academic.oup.com/biomet/article-abstract/26/4/404/291538>.
- [2] Peter Grünwald, Rianne de Heide, and Wouter M. Koolen. Safe testing. *Journal of the Royal Statistical Society Series B*, 86(5):1091–1128, 2024. doi: 10.1093/jrsssb/qkae011. URL <https://arxiv.org/abs/1906.07801>.
- [3] Steven R. Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform, nonparametric, nonasymptotic confidence sequences. *Annals of Statistics*, 49(2):1055–1080, 2021. doi: 10.1214/20-AOS1991. URL <https://arxiv.org/abs/1810.08240>.
- [4] Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *Journal of Machine Learning Research*, 17(1):1–42, 2016. URL <https://jmlr.org/papers/v17/kaufman16a.html>.
- [5] NIST/SEMATECH. e-handbook of statistical methods: Confidence intervals. Section 7.2.4.1, n.d. URL <https://www.itl.nist.gov/div898/handbook/prc/section2/prc241.htm>.
- [6] Jeremy Rodgers. Source-readout non-equivalence: Descent and equivariant reconstruction obstructions, 2026. URL <https://doi.org/10.5281/zenodo.21369967>.
- [7] Ian Waudby-Smith and Aaditya Ramdas. Estimating means of bounded random variables by betting. *Journal of the Royal Statistical Society Series B*, 86(1):1–27, 2024. doi: 10.1093/jrsssb/qkad009. URL <https://arxiv.org/abs/2010.09686>.