Compressed Records and the Queries They Can Answer: Forwarding Tables, Deadline Counters and Event Lists

Flyxion
Independent Researcher

November 2026

Abstract

Systems with limited memory, bandwidth and trust keep compressed records of what they observe and answer later questions from the records alone. Whether a record is adequate depends on which questions will be asked, which observations can occur, and how the answers will be checked, and claims of adequacy that omit these are routinely over-broad. This essay states the adequacy condition as sufficiency of a representation for a family of queries, with an exact form (every query factors through the representation) and an approximate form (a stated tolerance), and gives two executable tools: a search for a refuting pair of observations and a check of a proposed factoring function. The formulation is applied to forwarding tables, deadline counters on a wrapping clock, threshold-crossing event lists of synthetic signals, a recorded arrhythmia excerpt, a second recording with two simultaneous channels, a noise-stress database with annotated beats and graded real noise, and manipulations designed against the detectors. Exact results include a characterization of when residues of a wrapping clock decide validity and a bound on reclamation work for safe expiry. Measured results show that aggregation preserves lookup but not lookup after withdrawal, that event lists reproduce interval statistics at about one two-hundredth of the storage until in-band noise pollutes them silently, that timing between channels is a far more sensitive carrier of evidence than timing within one, and that a ramp spreads a delay below the rate a level detector can see. The conclusion distinguishes sufficiency of a record from the reliability of the evidence it holds, and gives a reporting standard for claims of either.

Keywords: representation sufficiency; query families; forwarding tables; deadline counters; event lists; timing evidence; noise robustness; adversarial manipulation

1 Introduction

A system that keeps a compressed record of what it has observed is committed to a claim about the future: that the record will be adequate for whatever is asked of it. A forwarding table that has been aggregated to save memory is adequate if every destination still resolves to the same next hop. A table of deadlines held in a sixteen-bit counter is adequate if every record is still read as valid or expired correctly. A list of beat times that replaces a thousand times as many samples is adequate if the interval statistics computed from it match those of the original. In each case the claim is about a relation between three things: the representation, the family of questions that will be put to it, and the class of observations from which it was made. Claims in practice often name only the first.

The relation has an old and precise form in statistics, where a statistic is sufficient for a parameter when it retains everything the sample says about the parameter [5, 11], and where comparison of experiments and approximate sufficiency supply the language for what is lost when the statistic is not sufficient [2, 13]. The same idea appears in the theory of state abstraction for sequential decision problems, where an abstraction is admissible for a purpose if it preserves the optimal behaviour [9, 14], and in the information bottleneck, where a representation is judged by what it retains about a named target [20]. The contribution here is narrow and practical. It fixes the formulation in a form that is directly checkable on finite domains, shows that the question that matters in engineering is usually the one the formulation makes explicit, namely which family of queries and which class of observations, and applies it to a sequence of concrete systems in which the answers are measured and not assumed. The formulation continues earlier work on projections that preserve the answer to a question and on sufficiency under later revision of policy [6, 7, 8], and the terminology of those essays is kept.

The measurements come from an executable notebook that accompanies this essay and is cited in the text as the notebook. It contains reference implementations of longest-prefix routing and bounded temporal state, synthetic and recorded signals, and the detectors and manipulations discussed in the later sections, and every figure quoted below was produced by it. Costs are counted in abstract operations and in bytes under stated word sizes, never in wall-clock time, since the timing of an interpreter says nothing about an embedded target. Where an experiment can fail, the notebook asserts the outcome the analysis predicts, and results that contradicted an expectation are reported as they came.

Section 2 states the formulation and four propositions. Section 3 treats forwarding tables and temporal state, where exact results are available. Section 4 treats event lists of synthetic signals. Section 5 repeats the measurements on recordings that were not chosen by the author. Section 6 introduces real noise at graded levels with ground truth, and manipulations designed against the detectors. Section 7 draws the reporting standard that the results support, and Section 8 states the limits of the evidence.

2 Formulation

Let X be a space of observations and $D \subseteq X$ the domain of observations that can actually occur. A representation is a map $P : D \to S$. A query is a function $f : D \to Y_f$, and a family of queries is a set $\mathcal{F}$ of such functions, possibly with different codomains.

Definition 1 (Sufficiency)**.** *The representation P is* sufficient for $\mathcal{F}$ on D *if for every $f \in \mathcal{F}$ there exists $g_f : S \to Y_f$ with $f(x) = g_f(P(x))$ for all $x \in D$. Given a metric d_f on each Y_f and a tolerance $\varepsilon \geq 0$, P is* ε-sufficient *if there exist g_f with $d_f(f(x), g_f(P(x))) \leq \varepsilon$ for all $x \in D$ and $f \in \mathcal{F}$.*

The dependence on D is part of the definition and is not a technicality: the same map P is routinely sufficient on a restricted domain and insufficient on a larger one. The dependence on $\mathcal{F}$ is equally essential.

Proposition 1 (Kernel criterion)**.** *P is sufficient for $\mathcal{F}$ on D if and only if for all $x, x' \in D$, $P(x) = P(x')$ implies $f(x) = f(x')$ for every $f \in \mathcal{F}$.*

Proof. If $f = g_f \circ P$ then equal images under P have equal values of f. Conversely, define $g_f(s) = f(x)$ for any $x \in D$ with $P(x) = s$, which is well defined by the hypothesis, and arbitrarily outside the image of P. □

Write $x \sim_{\mathcal{F}} x'$ when $f(x) = f(x')$ for all $f \in \mathcal{F}$. Proposition 1 says that P is sufficient exactly when its fibres lie inside the classes of $\sim_{\mathcal{F}}$. The coarsest sufficient representation is therefore the quotient $D/\!\sim_{\mathcal{F}}$, and any exactly sufficient representation that assigns every class a code of one fixed length needs at least $\lceil \log_2 |D/\!\sim_{\mathcal{F}}| \rceil$ bits. This is the minimal storage for the family under fixed-length coding, and it is a property of the family and the domain alone. It is not a bound for variable-length codes, probabilistic encodings or representations judged under a distribution over D, and the byte counts reported later are actual encodings and are not claimed to meet it.

The criterion yields two executable tools on a finite domain. A *refuting pair* is a pair x, x' with $P(x) = P(x')$ and $f(x) \neq f(x')$ for some $f \in \mathcal{F}$, and it shows by the proposition that no g_f exists; a search for one is a single pass over the domain with a hash of $P(x)$. A *certificate* is an explicit g_f whose defect, the number of $x \in D$ with $f(x) \neq g_f(P(x))$, is zero. In the approximate case the defect is replaced by the largest distance. Both are implemented in the notebook and used throughout.

Proposition 2 (Composition)**.** *If $P : D \to S$ is sufficient for $\mathcal{F}$ and $Q : S \to S'$ is sufficient on $P(D)$ for the family $\{g_f : f \in \mathcal{F}\}$, then $Q \circ P$ is sufficient for $\mathcal{F}$ on D.*

Proof. For $f \in \mathcal{F}$ take $g_f \circ h$, where h factors g_f through Q on $P(D)$. □

Composition licenses compression pipelines that are certified stage by stage, and it fails without the hypothesis on the second stage, which is the usual way in which a later compression step quietly removes what an earlier one retained.

The most consequential extension concerns systems that are updated. Let U be a set of maps $u : D \to D$ (insertions, withdrawals, renewals) and let $\mathcal{F}_U$ consist of all $f \circ u$ with $f \in \mathcal{F}$ and u a finite composition of elements of U, including the empty composition.

Proposition 3 (Closure under updates)**.** *P is sufficient for $\mathcal{F}_U$ on D if and only if for all $x, x' \in D$ with $P(x) = P(x')$ and every finite composition u of updates, $u(x) \sim_{\mathcal{F}} u(x')$.*

Proof. Apply Proposition 1 to $\mathcal{F}_U$. □

The condition is a congruence requirement of the same kind as in bisimulation and in model minimization [9]: the representation must identify only those observations that no future sequence of updates can separate. Sufficiency for the static family $\mathcal{F}$ does not imply it, and Section 3 gives a forwarding-table example in which the static family is preserved exactly and the family closed under withdrawals is not.

Section 3 also uses the following exact characterization for the representation that a wrapping clock provides. Let $M = 2^B$ and $h = M/2$. A record has deadline d and is read at time t, with remaining lifetime $\delta = d - t$ and validity valid $= [\delta > 0]$. The representation is $P(d, t) = (d \bmod M,\ t \bmod M)$, which is what a B-bit counter holds. Let $\sigma(z)$ denote the representative of $z \bmod M$ in $[-h, h-1]$.

Proposition 4 (Residues of a wrapping clock)**.** *Let $D_{a,b}$ be the set of pairs (d, t) with $\delta \in [-a, b]$, where $a \geq 0$ and $b \geq 1$ are integers. Then P is sufficient for* valid *on $D_{a,b}$ if and only if $a + b + 1 \leq M$. The signed comparison $g(u, v) = [\sigma(u - v) > 0]$ is a factoring function on $D_{a,b}$ if and only if $a \leq h$ and $b \leq h - 1$.*

Proof. The value of valid depends only on δ, and P determines $\delta \bmod M$ and nothing finer. If $a + b + 1 \leq M$ the integers in $[-a, b]$ have distinct residues, so δ is determined and the criterion holds. If $a + b + 1 > M$, put $\delta_0 = \min(0, b - M)$. Then $\delta_0 \geq -a$ because $a + b \geq M$, and $\delta_0 + M = \min(M, b)$ lies in $[1, b]$, so both δ_0 and $\delta_0 + M$ lie in the interval. They share a residue and differ in validity, which refutes sufficiency. For the signed comparison, $\sigma(\delta \bmod M) = \delta$ exactly when $\delta \in [-h, h-1]$; the value $\delta = h$ is read as $-h$ and is declared expired although valid, and $\delta = -h - 1$ is read as $h - 1$ and is declared valid although expired, so either endpoint beyond the window produces an error. □

The convention valid $= [\delta > 0]$ is strict: a record is expired at the tick of its deadline. The common comparison of elapsed time against a timeout, which declares a record live while elapsed time is at most the timeout, corresponds to $[\delta \geq 0]$, and shifts the endpoints of the windows in the proposition by one. The statement is therefore tied to the convention, and an implementation has to fix one and test the boundary ticks on both sides.

A fixed-capacity table of N slots is swept cyclically, k consecutive slots per tick, so that each slot is visited at regular intervals of $\lceil N/k \rceil$ or $\lfloor N/k \rfloor$ ticks, and let P denote the largest of these gaps. The execution model is sequential: within a tick the sweep runs first, then insertions, then lookups, with no concurrent mutation. A visit releases a slot whose signed remaining lifetime is

non-positive. An insertion writes only into a free slot, so an occupied record is never overwritten without having been resolved, and it assigns a lifetime of at least 1 and at most $h - 1$.

Proposition 5 (Reclamation bound)**.** *In this model, if $k \geq 1$ and $P \leq h + 1$ then no lookup ever reports a record valid after its deadline. If $P \geq h + 2$ there is an execution of the same model in which one does.*

Proof. Suppose $P \leq h + 1$ and a record has deadline d. Until d its signed remaining lifetime lies in $[1, h - 1]$ and it is read as valid correctly. The first visit of its slot at or after d occurs at some tick $d + j$ with $j \leq P - 1 \leq h$, where the signed value $-j \in [-h, 0]$ is non-positive and the record is released before any lookup of that tick, and a record not yet visited is read correctly because $-j \geq -h$. For the converse, take a slot whose gap is $P \geq h + 2$, which exists because the largest gap is attained, and insert into it immediately after one of its visits at tick t, with lifetime $\tau = \min(P - h - 1, h - 1) \geq 1$, so that $d = t + \tau$. The next visit is at $t + P \geq d + h + 1$, and no earlier visit exists. At tick $d + h + 1$ the signed value is $-(h + 1) \equiv h - 1$, which is read as valid; if that tick is itself the visit, the sweep keeps the record for the same reason, and the lookup that follows reports it valid after its deadline. □

The proposition turns reclamation delay into a correctness parameter: the sweep budget cannot be chosen from a latency target alone, and the counter width, the maximum lifetime, the capacity and the sweep budget have to be specified together.

3 Forwarding Tables, Counters and Deadlines

Longest-prefix matching is the standard setting for compact representations of a table that must answer a fixed query quickly [19, 23]. The notebook implements a linear reference, a binary trie and a length-indexed hash, and tests the latter two against the reference on 258,544 pairs of generated table and address, 76,800 of them exhaustive over an eight-bit address space and the remainder with probe addresses placed at the first and last address of every prefix and one step outside it, the places where an off-by-one error would show. No disagreement occurred. Measured in operations, the linear scan costs the table size, the trie visits between 17 and 22 nodes per lookup at sizes from 64 to 16,384 entries with a maximum of 33, and the length hash probes between 8 and 16 lengths on average. The trie therefore loses to the linear scan at 16 entries, where it averages 20 nodes, and its node count is several times the entry count (100,896 nodes for 16,384 entries). The length hash has the lowest counts at these sizes, although a hash probe is not a node visit and the two columns are not interchangeable. These are the usual trade-offs and are reported so that the later claims can be read against them.

The sufficiency analysis uses all $3^7 = 2187$ tables over a two-bit address space, with next hops in $\{\text{none}, A, B\}$ for each of the seven prefixes. Aggregation, defined as follows, preserves: a prefix is dropped when its next hop equals that of its longest strictly shorter covering prefix, with that covering prefix looked up in the original table and not in a progressively reduced one. Because a dropped prefix and its cover then resolve to the same next hop, whether or not the cover is itself dropped, the result does not depend on the order of removal. This is the operation

certified here, and it preserves every static lookup exactly: the defect against the lookup family is zero over all 2187 tables. It is not sufficient for the number of announced prefixes, refuted by the tables $\{1{:}A\}$ and $\{1{:}A,\ 11{:}A\}$, which aggregate to the same table. It is also not sufficient for the family of lookups after withdrawal of a covering prefix. The tables $\{0{:}A\}$ and $\{0{:}A,\ 01{:}A\}$ aggregate to the same table, and after withdrawal of the prefix 0 the first resolves every address to no route while the second still resolves the addresses beginning with 01 to A. This is the situation of Proposition 3: the static family is preserved and the family closed under withdrawals is not, because the aggregated form has forgotten that the more specific entry was redundant only while its parent existed. A second representation, the set of covered addresses, is sufficient for reachability and refuted for the next hop by the tables $\{11{:}A\}$ and $\{11{:}B\}$. Aggregation of forwarding tables has long been studied for its size savings [3]; the point here is the family against which such savings are to be certified.

For temporal state the notebook compares a fixed-capacity table with deadlines kept on a wrapping counter against an unbounded reference. Proposition 4 is confirmed by exhaustive enumeration for $B = 4$: the defect of the signed comparison is zero on remaining lifetimes in $[-8, 7]$ and positive (48 of the pairs enumerated) on $[-9, 7]$ and on $[-8, 8]$, and a refuting pair exists in each of the latter, for example $(d, t) = (1, 16)$ and $(17, 16)$, whose residues agree and whose validity differs. This is the formal content of the convention for serial number arithmetic [4], with the horizon stated and enforced at insertion. Proposition 5 was tested over a grid of capacities and sweep budgets at $B = 6$, where $h + 1 = 33$, under the execution model stated there. Every cell agreed with the bound, with no premature expiry observed: $(N, k) = (33, 1)$ produced no false-valid answer in 60,000 ticks and $(34, 1)$ produced 101, and a purely lazy table with $k = 0$ produced 1297 for a pool of 16 keys. Hashed and hierarchical timing wheels [21] are the standard structure for efficient expiry, and the proposition states what any such structure must additionally guarantee when its clock wraps.

A bound on allocated records is not a bound on useful service. The table was driven by a steady stream of legitimate records with lifetime 8 and, in a second phase, by an adversary inserting one long-lived record per tick, with capacity 64. Rejecting new records admitted 29 percent of the legitimate insertions during the flood, all of which lived out their lifetimes, because the long-lived flood occupied the slots. Evicting the record with the least remaining life admitted all of them and none survived to the end of its lifetime, because the short-lived legitimate records were always the cheapest to evict, and each insertion into the full table cost on average 79 slot touches per tick against 2 without the scan. A specification of bounded state is incomplete until it states which of these it chooses and what it guarantees to whom.

Two further representations of a deadline complete the picture. Rounding it up to a multiple of q, as a timing wheel does, is $(q - 1)$-sufficient for the expiry time and never expires a record early, as the notebook verifies for q from 1 to 16. Storing only the latest deadline is sufficient for validity at every time and refuted for the number of renewals by the histories (0) and $(0, 0)$.

4 Event Lists from Synthetic Signals

An event list replaces a signal by a short list of times at which something happens, here the upward zero crossings, each located by linear interpolation between the two samples that straddle zero. The synthetic signal is a ten hertz oscillation with a slow phase wobble and a weak third harmonic, observed at 2000 samples per second for twenty seconds, and the ground truth for crossing times comes from the noiseless signal on a grid fifty times finer.

On clean input the event list reproduces all 199 crossings with a root-mean-square error of 0.014 milliseconds when times are stored at 50 microsecond resolution, in 398 bytes against 80,000 for the sixteen-bit samples, a ratio of about 200. Sampling at 100 hertz gives an error of 0.11 milliseconds in 4000 bytes and at 50 hertz 0.82 milliseconds in 2000 bytes, so for this signal class and this query the event list is cheaper than any sampled representation that achieves a comparable error. With additive noise of standard deviation 0.05, a bare threshold chatters, producing 178 false events; hysteresis removes them at that level and, at noise 0.15, leaves 160. The error that remains is the noise itself converted to time by the slope at the crossing.

Linear interpolation is invariant under positive scaling, so a signal and the same signal multiplied by two have identical crossing times, and the family containing any amplitude query is refuted by that pair. The approximate version is measured by reconstruction: from crossing times alone the best generic reconstruction assumes unit amplitude, and from extremum events (a time and an interpolated value) it joins successive extrema by half-cosines. On the constant-amplitude signal the crossing list gives a root-mean-square error of 0.037 and the extremum list 0.081, so the crossing list is the better of the two and nearly sufficient for the waveform. With fifty percent amplitude modulation the crossing-only error rises to 0.254 while the extremum error stays at 0.085. Neither representation dominates, and which is sufficient, to what tolerance, depends on the class of signals.

For two channels, the phase-locking statistic of Lachaux and colleagues [12] depends on the second channel through its crossing times alone when phase is interpolated between crossings. Over settings from strong coupling to none, with twenty noise realizations each at noise 0.05, the event-based statistic stayed within 0.09 of the statistic computed from the full signals via the analytic signal, so the event list is approximately sufficient for the query at that tolerance on that class. The same property bounds what the statistic can certify. A signal synthesized from the stored event times, with a different amplitude and no noise, reproduces the event times to within 0.023 milliseconds, against a storage resolution of 0.050, and the same statistic of 0.962 as the original recording, while its amplitude (0.49 against 0.71) and its noise floor (none against 0.05) differ visibly. A label such as recorded versus synthesized cannot factor through the event list, since the two share it, and any provenance claim has to rest on evidence outside the representation.

The last synthetic query is the detection of an inserted delay. The score is the standardized change between successive intervals, with scale estimated robustly from the first one hundred events. A delay of 5 milliseconds was detected in 5 percent of trials and one of 10 milliseconds in all of them, with no false alarm in twenty unmodified trials, and the later sections show how much of that sharpness belongs to the signal class.

5 Recorded Signals

5.1 An arrhythmia excerpt

The first recording is the five-minute electrocardiogram excerpt distributed with scipy [22], sampled at 360 hertz and carrying no beat annotations, so the reference is the output of a detector in the lineage of Pan and Tompkins [18] run on the full-rate signal, and every figure below is agreement with that reference. The reference is itself uncertain: moving the detector threshold changes the beat count from 425 to 455 or 411. Because the interval statistics below are defined from the same detected beat times that the list stores, their preservation is nearly built into the experiment and shows that the storage is lossless for this query, and not that the detector is right; the independent annotations of Section 6 are the test of the second claim. Against the detector output, an event list of 425 beat times stored to one millisecond occupies 850 bytes, against 216,000 for the samples, and reproduces the mean heart rate, the standard deviation of intervals and the root mean square of successive differences to 0.001 beats per minute, 0.005 milliseconds and 0.02 milliseconds (88.353, 274.218 and 357.788 against 88.354, 274.223 and 357.808). Sampled forms agree exactly at 180 hertz and above, deviate by about one percent in the interval standard deviation at 120 hertz, and fail badly below: at 60 hertz 15 false beats raise the heart rate to 91.1 and lower the standard deviation to 242, and at 30 hertz 90 false beats give a heart rate of 103.

A clinician asks whether a beat looks like the others as well as when it occurred. The morphology label is defined from the recording by correlating each beat window with the median beat and calling a beat deviant below 0.8, which labels 53 of 415 beats. The candidate record uses the interval before and after each beat, normalized by the median of the preceding eight, and optionally the peak amplitude. The defect is measured as the cross-validated error of the best table-lookup predictor from the quantized stored fields, since no analytic factoring function is available. The baseline of always predicting the common label has error 0.128, timing alone has 0.130, and timing with amplitude has 0.104. In this recording timing alone does not predict morphology better than saying every beat is normal, and a refuting pair of beats with the same quantized timing context and correlations of 0.95 and 0.59 shows why; the amplitude field recovers about a dozen beats of 415, a gain worth stating and within what a different split could move.

The delay detector behaves very differently on this recording from the synthetic one. The natural variability of the intervals is so large that at a threshold of 6 the score already exceeds it at 27 percent of all interval changes, and the excess over the unmodified recording at the same positions is the quantity to read. A delay of 111 milliseconds produced no excess to speak of; 556 milliseconds produced an excess of 0.60 at threshold 6 and 0.14 at threshold 20, and 1112 milliseconds produced 0.66 and 0.93. The ten millisecond sensitivity of the synthetic case is the difference between a stable oscillator and a heart, and a detection margin reported without its signal class is misleading.

5.2 A regular rhythm and a second channel

Two recordings from the sample data of NeuroKit2 [15] separate the dependence on the class of signal from the dependence on one excerpt. The first is a fifty-second electrocardiogram at 1000 hertz with 50 beats and a regular rhythm. Sampled forms with the same detector agree with the full-rate reference without a missed or false beat down to 62.5 hertz, and fail at 31 hertz, with 12 missed and 12 false beats and an interval standard deviation of 71 milliseconds against 62. The sampling floor of the arrhythmia excerpt, between 120 and 180 hertz, therefore does not transfer: a floor belongs to the record. With 50 beats this recording indicates plausibility and carries little statistical weight.

The second is a five-minute resting recording at 100 hertz with an electrocardiogram, a photoplethysmogram taken at the pulse, and respiration. It contains 436 beats and 432 pulse peaks, and permits a query that no single-channel record can answer, the pulse transit time from beat to the first pulse peak within 0.2 to 0.7 seconds. The median is 434.6 milliseconds with an interquartile range of 15.2. Event lists preserve it at 100 and 50 hertz (median 433.6 at 50) and degrade at 25 hertz, where the beat detection collapses (121 beats missed and 135 false) even though the slowly varying pulse signal is tracked without error, and the interquartile range rises to 58.2 milliseconds. The failure is in one channel and the cross-channel statistic inherits it, so a specification of the representation for this query has to state a sampling floor per channel.

A delay inserted into the pulse channel only was detected by the transit time at a threshold of 6 with an excess over the natural rate of 0.80 at 40 milliseconds, while the interval structure of the pulse channel alone needed several hundred milliseconds (an excess of 0.98 at 320). The relation between two channels is about an order of magnitude steadier than either channel’s own intervals and carries correspondingly more sensitive evidence. When the same delay was inserted into both channels, the transit time did not change and the detector was blind, with an excess indistinguishable from zero. The cross-channel statistic detects an inconsistency between channels and is not evidence that a recording is unmodified.

6 Noise and Designed Manipulation

6.1 Graded noise with ground truth

The MIT-BIH Noise Stress Test Database [16], distributed through PhysioNet [10], provides the arrhythmia recording number 118 [17] with real recorded electrode-motion noise added at 24, 18, 12, 6, 0 and −6 decibels in two-minute stretches alternating with clean stretches, and every version shares the original beat annotations. The notebook verified the structure by subtracting the original from each noisy record: noise is present exactly in minutes 5 to 7, 9 to 11 and 13 to 15 of the fifteen minutes used, which contain 1146 annotated beats. A detection counts as correct within 150 milliseconds of an annotated beat, the usual tolerance [1]. Because the annotations do not depend on the detector, these are the first figures in the study that are agreement with an independent reference.

The clean record is detected perfectly at every rate down to 30 hertz, so for this beat shape

Table 1: Positive predictivity of beat detection at 360 hertz in the noisy stretches, for three noise types of equal total power added to the original recording. Electrode motion is reconstructed by the same procedure and differs from the noise in the database records, which was cut from a different excerpt.

Noise type	24 dB	12 dB	6 dB	0 dB
Electrode motion	1.00	0.82	0.59	0.50
Muscle artifact	1.00	0.86	0.70	0.58
Baseline wander	1.00	1.00	1.00	0.99

and rhythm the sampling floor lies below anything tried. Noise acts differently from sampling. Sensitivity stays at or above 0.93 down to 6 decibels and falls to 0.88 and 0.85 at 0 and −6, but positive predictivity falls from 1.00 at 24 and 18 decibels to 0.92 at 12, 0.69 at 6 and 0.54 at 0 decibels, so the event list fills with false beats and does not lose true ones. At a fixed noise level the sampling rate barely matters down to 60 hertz. Against ground truth, in the conditions tested (one record, one noise type at this stage, one detector and one set of thresholds), the event list matched the annotations without a false beat at 24 and 18 decibels and was visibly polluted from 12 decibels downward, and what separated the two regimes was the noise level and not the sampling rate. No universal boundary is claimed.

The pollution is silent and it propagates. Interval statistics computed from detections agree with those of the annotations to a few tenths of a millisecond down to 18 decibels (mean heart rate 77.2, standard deviation 80.8 and 109.9 for successive differences, in the units of the earlier sections), and at 12 decibels they read 84.3, 155.0 and 173.0, and at 6 decibels 106.9, 206.9 and 259.7. Nothing in the event list indicates that false beats have entered it, so a representation intended for interval queries needs a quality indicator stored beside the events, and in these data the transition lay between 18 and 12 decibels. The detection of an inserted delay degrades in the way the false beats predict. The excess detection of a 556 millisecond delay at threshold 20 falls from 0.90 on the clean record to 0.35 at 12 decibels and 0.10 at 6, while a delay of 1112 milliseconds is still detected with an excess of 0.95, 0.98 and 0.83. Noise therefore moves the smallest detectable delay from about half a second to above a second.

The database also provides noise-only recordings of muscle artifact and baseline wander. Added to the original at the same total power as the electrode-motion noise, they behave very differently (Table 1). Baseline wander leaves the event list intact because 99 percent of its power lies below 1 hertz and the detector's band-pass removes it. The ordering follows the noise inside the detector's band: the rank correlation of positive predictivity with in-band noise level is 0.96 over the twelve conditions, against 0.78 with total noise level, which cannot separate the three types because they share it. Part of this agreement is by construction, since the in-band level is measured through the detector's own filter, and the practical statement is that a signal-to-noise ratio has to be measured in the band the representation uses. A figure for the whole signal does not say whether an event list is polluted. The same electrode-motion noise cut from a different excerpt gave positive predictivity of 0.92 and 0.69 at 12 and 6 decibels in the database records, against 0.82 and 0.59 here, so the excerpt alone moves the figure by about 0.1 and the difference

Table 2: Detection rate in 26 quiet windows of 130 beats, by detector, manipulation and total delay in milliseconds. A ramp delays pulses by a shift growing linearly over 100 beats to the stated total. "Both" applies the ramp to both channels.

	Per-beat, threshold 40		Level, threshold 30 ms		
Total	Step	Ramp	Step	Ramp	Both
20	0.00	0.00	0.00	0.00	0.00
40	0.00	0.00	1.00	0.00	0.00
80	0.08	0.00	1.00	1.00	0.00
160	1.00	0.00	1.00	1.00	0.00

between electrode motion and muscle artifact is of that size, whereas the difference from baseline wander is not.

6.2 Manipulation designed against the detector

Every manipulation so far was drawn at random and was a single step. An adversary who knows that a detector scores the change between successive transit times would not insert a step. The cheapest design against such a score is a ramp, the same total delay spread over many beats so that the change at any one beat stays below the scale the detector expects. Two detectors were tested on the resting recording: the per-beat score of the previous section, and a level detector built for the attack, which compares the median transit time of the last 20 beats with that of the 100 before and whose statistic D_k is the difference.

The transit time of this recording contains natural outliers, a handful of beats where a pulse is missed or mismatched, which push the per-beat score to about 67 and the level statistic to about 26 milliseconds. The attack windows were therefore placed in the one stretch where the natural level statistic stays below 15 milliseconds, with 26 start positions, and the natural alarm rates over all windows are reported alongside. All figures are detection rates over those windows (Table 2). At the sensitive per-beat threshold of 20, the natural alarm rate is 1.00 in every window and the detector says nothing. At the conservative threshold of 40 it detects a step of 160 milliseconds in every window and a step of 80 milliseconds in 8 percent, and it never detects the same 160 milliseconds delivered as a ramp over 100 beats, because the change at any beat is about 1.6 milliseconds. The level detector does better but does not close the attack. At a threshold of 30 milliseconds, which has no natural alarms anywhere in the recording, it detects a step of 40 milliseconds and a ramp of 80, and misses a ramp of 40, so the ramp costs the adversary a factor of two in total shift. At a threshold of 15 milliseconds it detects a ramp of 40 milliseconds, but that threshold alarms in 84 percent of windows of the unmodified recording. A ramp applied to both channels was detected in no window at any shift or threshold.

The behaviour of the level detector under sustained ramps follows from its definition and is stated as a bound.

Proposition 6 (Blind rate of a level detector)**.** *Let the statistic compare the median of the last r values with that of the preceding m values of a sequence, and suppose the sequence is a constant*

plus ε per step beyond some start, once both windows lie within the ramp. The statistic equals $\varepsilon(r+m)/2$. A ramp therefore stays below a threshold τ if $\varepsilon < 2\tau/(r+m)$, whatever its duration, in the absence of natural variation.

Proof. The median of a linear sequence over a window is its value at the centre of the window, and the centres of the two windows are $(r+m)/2$ steps apart. □

With $r = 20$ and $m = 100$ the blind rate is $\tau/60$ per beat. A sustained ramp, evaluated up to beat 265 so that every alarm in the stretch came from the attack (the natural level statistic never exceeds 11.4 milliseconds there), confirms it. At a threshold of 30 milliseconds no run was detected at 0.4 milliseconds per beat or below, over a stretch long enough for the ramp to reach 53 milliseconds, while 58 percent were detected at 0.5 and all at 0.6. At a threshold of 15 milliseconds the transition lay between 0.2 and 0.3. Once above the blind rate the alarm came at an accumulated shift of 41 to 46 milliseconds at the higher threshold and 18 to 27 at the lower, so the adversary obtains roughly the threshold's worth of shift before an alarm in either case. The proposition describes the steady state of a ramp: at its onset and at its end, where the windows straddle the change of slope, the statistic differs and the geometry has to be treated separately. A detector that compares against a fixed baseline authenticated at an earlier time does not have the same unbounded drift, since its statistic grows with the total shift, although it inherits the questions of how the baseline is validated and how far the natural level drifts over the period. The statistic does not depend on duration, so over a recording long enough to show it the undetected total would grow in proportion to time; the recording here is too short to test that extrapolation, and it is stated as following from the proposition.

Sufficiency and reliability

The propositions above concern a representation of the observations that were recorded. A record can be exactly sufficient for a query defined on its own events and still be unreliable evidence about the world, because the events were produced by a detector that can be wrong. Let Z be the physical state, $O : Z \to D$ the observation process together with the detector, $P : D \to S$ the representation, and $h : Z \to Y$ a query about the world. Sufficiency of P for a record-level query f on D says that f factors through P; it says nothing about $f \circ O$ against h. Reliability is the separate requirement that $d_Y\big(h(z), g(P(O(z)))\big) \le \varepsilon$ over the states z that can occur, and it fails when O inserts false events or drops true ones, however exact the factoring through P. The noise experiments are an instance: the event list preserved the interval statistics of the detected beats exactly while the detected beats stopped corresponding to the annotated ones. The distinction fits the reporting standard below, whose third and fourth items are exactly the reference and the noise that separate the two properties. Preserving all the information needed to answer a question is not the same as having reliable evidence that the answer is true.

7 A Reporting Standard

The results support a standard for claims about compressed records, which can be stated as five items that every claim should name. The first is the family of queries, including the closure of the family under updates when the record is maintained (Proposition 3). Aggregation of a forwarding table is sufficient for lookup and not for lookup after withdrawal, and an event list is sufficient for interval statistics and not for morphology or origin. The second is the domain of observations. The wrapping counter is sufficient on a window of M remaining lifetimes and not outside it (Proposition 4), and a sampling floor of 180 hertz on one record is a floor of 62.5 hertz on another. The third is the reference against which agreement is measured, which is itself an experimental variable: the threshold alone moved the beat count of the arrhythmia excerpt by seven percent, and only the annotated database gave a reference independent of the detector.

The fourth is the noise level measured in the band the representation uses, together with a stored indicator when the representation can silently fail. In the conditions tested the event list contained no false beats at 24 and 18 decibels and filled with them from 12 decibels downward, with no sign in the record. The fifth is the adversary. A detection margin is a function of the manipulation as well as of the detector: the same 160 milliseconds was invisible to the per-beat score as a ramp and obvious as a step, a level detector bounded the rate of movement and not the total, and cross-channel statistics were blind to a manipulation applied consistently to both channels. A margin quoted without the manipulation is not a margin.

A sixth consideration is not a fifth item but a separation. None of the timing statistics here bears on origin. The forged channel reproduced the event times and the phase-locking statistic, and a ramp on both channels left the transit time unchanged, so evidence of origin has to be carried by authenticated records and not by the signal, and signal analysis supplies at most a fallible measurement alongside.

8 Limits of the Evidence

The notebook is the file `temporal-systems.ipynb`. Its random generators are seeded, its dependencies are numpy, scipy, wfdb and NeuroKit2, and its recordings are the electrocardiogram distributed with scipy, the NeuroKit2 sample data and PhysioNet record 118 with its noise-stress variants, all fetched at run time and not bundled. Its assertions are of two kinds that should not be confused: some check mathematical invariants and are expected to hold for any seed, such as agreement of the routing structures with the reference and the exact window of the signed comparison, while others record an empirical outcome observed at the stated seed, such as a detection rate, and may change with a different excerpt or threshold. A stable repository path and revision identifier are needed before the figures can be cited as reproducible and are to be added with publication of the notebook.

The synthetic signals describe the behaviour of the representations and not of any real source. The recordings are single excerpts or small sets, one detector was used throughout, and the beat-based findings have been tested only on electrocardiograms and one photoplethysmogram;

different detectors, different morphologies and different physiology would move the numbers. The adversarial results use 26 overlapping windows from one recording, with thresholds whose natural alarm rates are in-sample figures, and indicate where each detector starts to work and not calibrated probabilities. The exact propositions in Section 2 are proved; the measured claims are measurements of a Python implementation, and the cost comparisons are in operations and bytes and carry no claim about any hardware. The sustained-ramp extrapolation over longer recordings and the effect of longer baseline windows on the blind rate are not tested.

9 Conclusion

A compressed representation is adequate for a family of questions over a domain of observations, to a stated tolerance, against a stated reference, under stated noise and stated adversaries. The formulation of sufficiency makes each of these a checkable parameter, and its two tools, a refuting pair and a certificate, turn most of the claims above into short computations. The measured picture is consistent across systems that have nothing else in common: representations that preserve one family of queries exactly are routinely insufficient for a neighbouring family, the neighbouring family is usually the one closed under updates or manipulated by an adversary, and the evidence a representation can carry about timing is not evidence about origin. The natural continuations are a longer recording to test sustained ramps against detectors with longer baselines, a second beat detector and a second class of signals, and the interface between event lists and authenticated records, where the quality indicator and the provenance claim belong.

References

[1] ANSI/AAMI. *Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms*. Standard EC57:2012. Association for the Advancement of Medical Instrumentation, 2012.

[2] D. Blackwell. Equivalent Comparisons of Experiments. *Annals of Mathematical Statistics*, 24(2):265–272, 1953.

[3] R. P. Draves, C. King, S. Venkatachary and B. D. Zill. Constructing Optimal IP Routing Tables. In *Proceedings of IEEE INFOCOM*, pages 88–97, 1999.

[4] R. Elz and R. Bush. Serial Number Arithmetic. RFC 1982, Internet Engineering Task Force, 1996.

[5] R. A. Fisher. On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society of London A*, 222:309–368, 1922.

[6] Flyxion. Lossless Projection. `https://standardgalactic.github.io/laboratory/lossless-projection.pdf`, 2026.

[7] Flyxion. Question-Relative Sufficiency in Learned World Models: From Safe Control to Independent Audit. Unpublished essay, 2026.

[8] Flyxion. Sufficient Projections. `https://standardgalactic.github.io/antivenom/economics/sufficient_projections.pdf`, 2026.

[9] R. Givan, T. Dean and M. Greig. Equivalence Notions and Model Minimization in Markov Decision Processes. *Artificial Intelligence*, 147(1–2):163–223, 2003.

[10] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng and H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation*, 101(23):e215–e220, 2000.

[11] P. R. Halmos and L. J. Savage. Application of the Radon-Nikodym Theorem to the Theory of Sufficient Statistics. *Annals of Mathematical Statistics*, 20(2):225–241, 1949.

[12] J.-P. Lachaux, E. Rodriguez, J. Martinerie and F. J. Varela. Measuring Phase Synchrony in Brain Signals. *Human Brain Mapping*, 8(4):194–208, 1999.

[13] L. Le Cam. Sufficiency and Approximate Sufficiency. *Annals of Mathematical Statistics*, 35(4):1419–1455, 1964.

[14] L. Li, T. J. Walsh and M. L. Littman. Towards a Unified Theory of State Abstraction for MDPs. In *Proceedings of the Ninth International Symposium on Artificial Intelligence and Mathematics*, 2006.

[15] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel and S. H. A. Chen. NeuroKit2: A Python Toolbox for Neurophysiological Signal Processing. *Behavior Research Methods*, 53(4):1689–1696, 2021.

[16] G. B. Moody, W. E. Muldrow and R. G. Mark. A Noise Stress Test for Arrhythmia Detectors. *Computers in Cardiology*, 11:381–384, 1984.

[17] G. B. Moody and R. G. Mark. The Impact of the MIT-BIH Arrhythmia Database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3):45–50, 2001.

[18] J. Pan and W. J. Tompkins. A Real-Time QRS Detection Algorithm. *IEEE Transactions on Biomedical Engineering*, 32(3):230–236, 1985.

[19] K. Sklower. A Tree-Based Packet Routing Table for Berkeley Unix. In *Proceedings of the USENIX Winter Conference*, pages 93–99, 1991.

[20] N. Tishby, F. C. Pereira and W. Bialek. The Information Bottleneck Method. In *Proceedings of the 37th Annual Allerton Conference on Communication, Control and Computing*, pages 368–377, 1999.

[21] G. Varghese and T. Lauck. Hashed and Hierarchical Timing Wheels: Data Structures for the Efficient Implementation of a Timer Facility. In *Proceedings of the Eleventh ACM Symposium on Operating Systems Principles*, pages 25–38, 1987.

[22] P. Virtanen et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[23] M. Waldvogel, G. Varghese, J. Turner and B. Plattner. Scalable High Speed IP Routing Lookups. In *Proceedings of ACM SIGCOMM*, pages 25–36, 1997.