Non-Identifiability of Polynomial Token Components

Exact, Approximate, and Repeated-Release Reconstruction Limits

Quantitative Reconstruction Limits Edition

Polygen Technical Note

2 October 2026

Abstract

This paper establishes precise conditions under which the coefficient component or the residual component of a least-squares polynomial token fails to identify the original data. For a fixed basis matrix $B \in \mathbb{R}^{L\times K}$ of rank r, the coefficient observation $C = B^{+}Y$ has fibers of dimension $(L-r)D$, while the residual observation $R = (I - BB^{+})Y$ has fibers of dimension rD, where $Y \in \mathbb{R}^{L\times D}$ and B^{+} denotes the Moore–Penrose pseudoinverse. Consequently, coefficients alone admit no universal exact inverse when $r < L$, and residuals alone admit no universal exact inverse when $r > 0$. We extend the result to constrained input domains and repeated fixed linear releases, whose joint invisible space is the intersection of the individual kernels. We also formalize policy-selected disclosure and split custody: complete observer views are ordered by deterministic refinement, added channels can only shrink or preserve compatible-input fibers and cannot worsen optimal reconstruction performance, and each coalition's fiber is the intersection induced by the channels available to that coalition. A designated combination may therefore be reconstruction-complete even when each separately administered downstream view is non-injective. We prove a separated-fiber bound for approximate and list reconstruction, derive Bayes-optimal exact and squared-error recovery under declared priors, and obtain closed-form posterior error and chi-squared laws under an isotropic Gaussian prior. Under a uniform finite prior, optimal exact recovery equals the number of attainable observations divided by the number of admissible inputs. Exhaustive enumeration of three declared single-reference component views gives exact success probabilities of 3.100586%, 2.443023%, and 6.909959%. For three selected fully enumerated fibers, minimum candidate separations of 2, 2, and $\sqrt{5}$ bound approximate-recovery success at error strictly below half the respective minimum separation by 1/55, 1/105, and 1/24 under their uniform conditional posteriors. If the complete released view is constant, the matched no-component baseline is $1/N$ and contributes zero mutual information. By contrast, a configuration-scoped replay that removes both numerical arrays but retains a prediction digest and serialized residual-magnitude field gives exact-recovery optima of 43.807983%, 64.526568%, and 69.191959% on the same finite domains. The latter is a decoded-metadata equation replay, not an operational package or natural-population result. We also give sharp deterministic boundaries for perfect secrecy and differential privacy. These are view-, metric-, prior-, and coalition-conditional reconstruction results, not cryptographic secret-sharing or confidential smart-contract guarantees and not evidence of practical inference resistance on an undeclared natural population.

Keywords: least squares, orthogonal projection, non-identifiability, polynomial tokenization, selected disclosure, split custody, coalition views, coefficients, residuals, metadata, reconstruction, privacy

---

Claim boundary. On the unrestricted fitting domain, coefficients have no universal exact inverse when $r < L$, and residuals have no universal exact inverse when $r > 0$. On a restricted domain, the same conclusion holds exactly when an attainable fiber is non-singleton. For a finite declared domain and prior, Theorem 5.2 gives the deterministic-view optimum and Remark 5.3 gives the general jointly distributed-view form. Corollary 5.8 applies only when the complete cumulative released view is constant on the prior support; withholding both numerical arrays does not make data-dependent metadata constant. The component censuses concern declared single-reference ideal views, while the metadata-only census is a separately scoped numerical-path replay. Every claim about a deployed view must include its ordered release history, metadata, side information, and other observables. Policy-selected disclosure means controlling the released observation map and measuring its consequences. It is a partial order, not a universal scalar security level: nested views are comparable, while coefficient-only and residual-only views are generally incomparable and require separate analysis. Split-custody claims are coalition-specific and must declare each party's holdings, permissible collusion, public protocol transcript, and any separate cryptographic protection. For repeated fixed linear releases of the same fitting-domain input, Theorem 4.2 computes the joint ambiguity from the intersection of their kernels; adaptive or nonlinear releases require the full joint observation map. The approximate-recovery bounds apply only to the stated metric,

radius, separated candidate set, and conditional prior. The Gaussian formulas additionally require a fixed observation operator and the declared isotropic Gaussian prior. The reported probabilities and information measures are not cryptographic security levels and do not establish practical inference resistance on representative natural populations, computational one-wayness, perfect secrecy, differential privacy, secure custody, confidential smart-contract execution, or package-level no-extra-leakage factorization. The paper does prove negative boundaries showing why deterministic fiber multiplicity alone cannot supply the latter guarantees. Build-specific claims must also satisfy the controlled conformance obligations in Appendix B.

1 Introduction

Polygen decomposes a numerical data segment into a fitted component and a residual component. In its simplest form,

$$Y = BC + R, \tag{1}$$

where B is a basis matrix, C contains the fitted coefficients, and R is the leftover error. Exact reconstruction in the fitting domain requires the sum of both components. That algebra alone, however, is not yet an impossibility proof. A defensible proof must identify the observation map, its domain, its kernel, the metadata available to the observer, and the conditions under which two admissible originals produce the same observation.

The phrase "impossible to reverse" is particularly easy to overstate. It can refer to at least six different claims:

1. **No universal exact inverse.** No decoder recovers the exact original for every admissible input.
2. **Prior- and loss-conditional reconstruction.** Exact, approximate, or list-recovery performance can be computed or bounded under a declared prior, metric, and complete observation.
3. **General inference resistance.** An observer cannot accurately approximate the input, identify a subject, or infer a sensitive attribute using priors and side information.
4. **Computational one-wayness.** Recovering an original is computationally infeasible.
5. **Perfect secrecy.** The released component reveals no information about the original.
6. **Differential privacy.** The release bounds how much any one protected contribution can change its output distribution.

This paper proves the first claim under explicit conditions and analyzes the second through exact finite-prior recovery, separated-fiber approximation bounds, Gaussian-prior squared-error laws, and fixed-linear repeated-release composition. It does not establish general inference resistance on an undeclared population, computational one-wayness, perfect secrecy, or differential privacy; it instead proves sharp negative boundaries for deterministic secrecy and differential privacy. Its central object is the *fiber*, the complete set of originals compatible with an observation. A component is exactly reversible on a specified domain if and only if every attainable fiber is a singleton. On the unrestricted fitting domain, producing a candidate is easy when the basis is known: coefficients yield BC, and residuals yield the zero-baseline candidate R. The obstacle is uniqueness. Once a domain, prior, metric, and complete view are declared, the fiber partition can also determine exact and approximate Bayes recovery and standard information measures, but not search cost or the accuracy of learned attacks on a different population.

2 Reference Model and Threat Model

2.1 Least-squares representation

Consider one segment in the fitting domain. Let

$$Y \in \mathbb{R}^{L\times D}, \qquad B \in \mathbb{R}^{L\times K}, \tag{2}$$

where L is the segment length, D is the number of data columns, and K is the basis width. Let

$$r = \operatorname{rank}(B) \tag{3}$$

and let B^+ denote the Moore–Penrose pseudoinverse. Define

$$P = BB^+, \qquad Q = I_L - P. \tag{4}$$

The matrix P is the orthogonal projector onto $\mathrm{col}(B)$, while Q is the orthogonal projector onto $\mathrm{col}(B)^\perp$ [1, 2]. The ideal least-squares representation is

$$C = B^+Y, \qquad R = QY, \qquad Y = BC + R. \tag{5}$$

The fitted component $BC = PY$ lies in $\mathrm{col}(B)$, and the residual satisfies

$$B^T R = 0. \tag{6}$$

The full-column-rank case has $r = K$ and is common in least-squares implementations. The theorems below use r so that their mathematical scope remains correct even when B is rank deficient.

2.2 Observers, roles, and coalitions

We distinguish five observers:

- A **no-component observer** receives neither numerical array. This is a constant-view observer only if every remaining in-scope field is fixed independently of the source.
- A **metadata-only observer** receives the complete structural and data-dependent transcript remaining after both numerical arrays are removed.
- A **component-only observer** receives C or R and nothing else.
- A **generous observer** receives one component, B, all structural metadata, and all inverse-preprocessing parameters. Proving non-uniqueness for this observer is stronger because it grants information that a real protocol may withhold.
- A **side-informed observer** additionally knows bounds, physical constraints, population statistics, training data, or auxiliary records.

The core linear results apply to the generous observer on an unrestricted real-valued domain. Restricted domains, side information, and prior-conditional recovery are handled separately in Sections 4 and 5.

The observer labels describe information sets rather than deployment roles. For a split-custody deployment, let

$$\mathcal{P} = \{\text{originator}, \text{custodian}, \text{receiver}\}. \tag{7}$$

The originator is presumed to know the source data when the representation is created unless a separate deletion or erasure claim is proved. The custodian may hold a complementary component, encrypted artifact, or release authority. The receiver obtains the selected task view. For every coalition $\Gamma \subseteq \mathcal{P}$, the relevant observation is the complete cumulative joint view of its members, including metadata, earlier releases, protocol messages, and side information available to that coalition.

If the custodian role is implemented by a public smart contract, plaintext state, calldata, events, transaction fields, return values, and published keys belong to the public view. A contract can enforce public release logic or coordinate encrypted material, but it does not create confidentiality by itself.

3 Exact Component-Fiber Theorem

Theorem 3.1 (Coefficient and residual fibers)**.** *Let $B \in \mathbb{R}^{L\times K}$ have rank r, and apply the observation maps in Equation (5) independently to each of D columns.*

For every attainable coefficient matrix C,

$$\left\{Y \in \mathbb{R}^{L\times D} : B^+Y = C\right\} = BC + \left\{Z \in \mathbb{R}^{L\times D} : B^TZ = 0\right\}. \tag{8}$$

Table 1: Illustrative selective-disclosure and split-custody roles. The exact holdings are policy choices; every role and coalition must be evaluated through its complete cumulative view.

Role	Possible material or authority	Scope of the resulting claim
Data originator	Source data initially, disclosure policy, and optionally inverse-preprocessing metadata or a release credential	Controls derivation and authorization, but is presumed to know the original absent a separate erasure claim
Custodian	A withheld component, encrypted artifact, wrapped key, or release authority	Its holdings form an observer view; trust, compromise, and collusion must be declared
Smart contract	Public policy state, commitments, encrypted references, attestations, and release events	Can coordinate and audit release; public state and transaction data provide no confidentiality by themselves
Receiver	The selected component and task-required metadata	Reconstruction and inference limits apply only to the receiver's complete cumulative view
Designated coalition	Combined components, basis information, and any source-inversion metadata used by the intended workflow	Exact recovery occurs only when the joint observation determines the intended target on the declared domain

This affine fiber has dimension $(L-r)D$.

For every attainable residual matrix R,

$$\left\{Y \in \mathbb{R}^{L\times D} : QY = R\right\} = R + \operatorname{col}(B)^D \tag{9}$$

$$= \left\{R + BA : A \in \mathbb{R}^{K\times D}\right\}. \tag{10}$$

This affine fiber has dimension rD.

Proof. Suppose $B^+Y = C$. Since $P + Q = I_L$,

$$Y = PY + QY = BB^+Y + QY = BC + QY. \tag{11}$$

The term QY lies in $\operatorname{col}(B)^\perp = \ker(B^T)$. Conversely, if $Y = BC + Z$ with $B^TZ = 0$, then $B^+Z = 0$. Because C is attainable, $B^+BC = C$, so $B^+Y = C$. Rank-nullity gives $\dim\ker(B^T) = L - r$ for each column, yielding dimension $(L-r)D$.

For residuals, $QY = R$ if and only if $Q(Y - R) = 0$. The null space of Q is $\operatorname{col}(B)$, so $Y - R$ lies in $\operatorname{col}(B)^D$. This proves Equation (10). Since $\dim\operatorname{col}(B) = r$, the fiber has dimension rD. □

Corollary 3.2 (Exact injectivity conditions)**.** *On* $\mathbb{R}^{L\times D}$*:*

- *coefficients alone are non-injective if and only if* $r < L$*;*
- *residuals alone are non-injective if and only if* $r > 0$.

For a full-column-rank compressed basis with $r = K < L$*, coefficient fibers have dimension* $(L-K)D$ *and residual fibers have dimension* KD.

Table 2 summarizes the complementary ambiguity.

The table exposes an important asymmetry. When $K \ll L$, coefficients are compact and omit most fitting-domain degrees of freedom. Residuals retain the entire orthogonal component and may therefore preserve much of the sample-level structure. A residual-only protocol can prevent exact reconstruction while still leaking a great deal. Those statements are compatible.

4 Restricted Domains and Side Information

Theorem 3.1 operates on an unrestricted real-valued space. Real applications may impose integer values, bounds, monotonicity, sparsity, physical laws, known statistics, or population priors. Let

$$\mathcal{S} \subseteq \mathbb{R}^{L\times D} \tag{12}$$

Table 2: Dimensions of the observed and unobserved subspaces for a full-column-rank basis. The final row concerns only the ideal fitting domain; source-domain recovery also requires the inverse-preprocessing information discussed in Section 9.

Released component	Observed degrees of freedom	Unobserved degrees of freedom	Exact non-injectivity condition
Coefficients C	KD	$(L-K)D$	$K < L$
Residuals R	$(L-K)D$	KD	$K > 0$
Both, with B known	LD	0	None for fitting-domain Y

be the actual admissible set after incorporating every constraint available to the observer.

Theorem 4.1 (Restricted-domain identifiability)**.** *The coefficient map is injective on $\mathcal{S}$ if and only if*

$$(\mathcal{S} - \mathcal{S}) \cap \ker(B^T)^D = \{0\}. \tag{13}$$

The residual map is injective on $\mathcal{S}$ if and only if

$$(\mathcal{S} - \mathcal{S}) \cap \operatorname{col}(B)^D = \{0\}. \tag{14}$$

Proof. Two inputs have the same coefficients exactly when their difference lies in $\ker(B^+) = \ker(B^T)$. They have the same residual exactly when their difference lies in $\ker(Q) = \operatorname{col}(B)$. Restricting each statement to differences between elements of $\mathcal{S}$ gives Equations (13) and (14). □

For a particular observed coefficient matrix C, define

$$\mathcal{F}_C(C) = \mathcal{S} \cap \left(BC + \ker(B^T)^D\right). \tag{15}$$

For a particular residual r_0, define

$$\mathcal{F}_R(r_0) = \mathcal{S} \cap \left(r_0 + \operatorname{col}(B)^D\right). \tag{16}$$

Exact reconstruction of that observation is possible if and only if its fiber is a singleton.

This qualification is not cosmetic. A bounded or highly structured domain can make a particular observation unique. With a mean-only basis and $Y \in \{0,1\}^2$, for example, an observed mean of zero identifies $(0,0)$ uniquely. The correct question is never merely "Is the map dimension reducing?" It is "How many admissible inputs remain in this specific fiber after every disclosed constraint is applied?"

4.1 Joint and repeated linear releases

The next result applies to fixed linear release maps acting on the same fitting-domain input. The matrices must not encode an additional undisclosed data-dependent selection event.

Theorem 4.2 (Kernel intersection for cumulative linear releases)**.** *For $t = 1, \ldots, T$, let*

$$A_t \in \mathbb{R}^{m_t \times L}, \qquad O_t(Y) = A_t Y, \tag{17}$$

where the matrices A_t are fixed independently of Y. Define the stacked operator

$$A_{1:T} = \begin{bmatrix} A_1 \\ \vdots \\ A_T \end{bmatrix}. \tag{18}$$

For every attainable cumulative observation $v_{1:T} = (v_1, \ldots, v_T)$ and any compatible Y_0,

$$\{Y : A_t Y = v_t \text{ for every } t\} = Y_0 + \left(\bigcap_{t=1}^{T} \ker A_t\right)^D \tag{19}$$

$$= Y_0 + \ker(A_{1:T})^D. \tag{20}$$

The unrestricted joint fiber has dimension

$$D\left(L - \operatorname{rank}(A_{1:T})\right). \tag{21}$$

Thus the cumulative release is injective on $\mathbb{R}^{L\times D}$ if and only if

$$\bigcap_{t=1}^{T} \ker A_t = \{0\}. \tag{22}$$

On a restricted domain $\mathcal{S}$, it is injective if and only if

$$(\mathcal{S} - \mathcal{S}) \cap \ker(A_{1:T})^D = \{0\}. \tag{23}$$

Finally,

$$\ker(A_{1:T+1}) = \ker(A_{1:T}) \cap \ker(A_{T+1}), \tag{24}$$

so cumulative invisible subspaces can only shrink as releases are added.

Proof. An input Y has the same cumulative observation as Y_0 if and only if

$$A_t(Y - Y_0) = 0 \qquad \text{for every } t. \tag{25}$$

This is equivalent to every column of $Y - Y_0$ belonging to $\bigcap_t \ker A_t = \ker A_{1:T}$, which proves Equation (20). Rank-nullity gives Equation (21). The unrestricted and restricted injectivity statements follow by setting the corresponding difference set equal to $\{0\}$. Equation (24) follows directly from the definition of the stacked operator. □

Corollary 4.3 (Repeated coefficient and residual releases)**.** *Let B_t be fixed basis matrices and $Q_t = I - B_t B_t^+$. For coefficient releases $C_t = B_t^+ Y$, the cumulative invisible subspace is*

$$\bigcap_{t=1}^{T} \ker(B_t^+) = \bigcap_{t=1}^{T} \operatorname{col}(B_t)^\perp = \left(\sum_{t=1}^{T} \operatorname{col}(B_t)\right)^\perp. \tag{26}$$

Writing

$$\rho_C = \operatorname{rank}\begin{bmatrix} B_1 & \cdots & B_T \end{bmatrix}, \tag{27}$$

the coefficient-history fiber on the unrestricted fitting domain has dimension $D(L - \rho_C)$ and is injective there exactly when $\rho_C = L$.

For residual releases $R_t = Q_t Y$, the cumulative invisible subspace is

$$\bigcap_{t=1}^{T} \ker Q_t = \bigcap_{t=1}^{T} \operatorname{col}(B_t), \tag{28}$$

and its fiber dimension is

$$D \dim \left(\bigcap_{t=1}^{T} \operatorname{col}(B_t)\right). \tag{29}$$

For a mixed history with coefficient indices $\mathcal{I}_C$ and residual indices $\mathcal{I}_R$, the invisible subspace is

$$\left(\sum_{t\in\mathcal{I}_C} \operatorname{col}(B_t)\right)^\perp \cap \bigcap_{t\in\mathcal{I}_R} \operatorname{col}(B_t). \tag{30}$$

In particular, releasing both $C = B^+Y$ and $R = QY$ for the same known basis leaves the zero invisible subspace and recovers $Y = BC + R$.

Proof. Use $\ker(B_t^+) = \ker(B_t^T) = \operatorname{col}(B_t)^\perp$, $\ker(Q_t) = \operatorname{col}(B_t)$, and the identity $\bigcap_t U_t^\perp = (\sum_t U_t)^\perp$, then apply Theorem 4.2. □

The kernel formulas require fixed linear maps. Quantization, changing preprocessing, data-selected bases, adaptive segmentation, and selection transcripts require analysis of the complete joint observation map. Joint fibers are always intersections of single-release fibers, but those fibers need not be affine kernels in an adaptive or nonlinear pipeline.

5 Prior-Conditional Exact Recovery and Information Leakage

Non-injectivity is a worst-case property: it establishes that some distinct admissible inputs have the same observation. A prior on a finite admissible domain supports a different question: what is the greatest exact-recovery probability achievable by any decoder that knows both the prior and the complete observation map?

Let $\mathcal{S}$ be a finite declared domain, let Y be an $\mathcal{S}$-valued random variable with prior

$$\pi(y) = \Pr[Y = y], \tag{31}$$

and let

$$V = O(Y) \tag{32}$$

denote the observer's complete cumulative deterministic view. The map O must include every released field and deterministic metadata item covered by the claim, together with every correlated earlier release and any disclosed certificate or transcript. Fixed side information can be folded into the domain or map; arbitrary correlated random side information is handled below. Define the positive-probability support

$$\mathcal{S}_\pi = \{y \in \mathcal{S} : \pi(y) > 0\}. \tag{33}$$

For every attainable observation $o \in O(\mathcal{S}_\pi)$, define

$$\mathcal{F}(o) = \{y \in \mathcal{S}_\pi : O(y) = o\}, \tag{34}$$

$$p_o = \Pr[V = o] = \sum_{y \in \mathcal{F}(o)} \pi(y), \tag{35}$$

$$\pi_o(y) = \Pr[Y = y \mid V = o] = \frac{\pi(y)}{p_o}, \qquad y \in \mathcal{F}(o). \tag{36}$$

5.1 Bayes-optimal exact recovery

Definition 5.1 (Optimal exact-recovery probability)**.** The optimal exact-recovery probability for Y given V is

$$P_{\text{guess}}(Y \mid V) = \sup_g \Pr[g(V) = Y], \tag{37}$$

where the supremum ranges over all deterministic or randomized decoders.

Theorem 5.2 (Bayes-optimal exact recovery)**.** *For every finite prior π and deterministic observation map O,*

$$P_{\text{guess}}(Y \mid V) = \sum_{o \in O(\mathcal{S}_\pi)} \max_{y \in \mathcal{F}(o)} \pi(y) \tag{38}$$

$$= \sum_{o \in O(\mathcal{S}_\pi)} p_o \max_{y \in \mathcal{F}(o)} \pi_o(y). \tag{39}$$

A maximum-a-posteriori decoder attains this probability. Randomization cannot improve it.

Proof. For each observation o, let $q_o(y)$ be the probability that a possibly randomized decoder outputs y. Its exact-recovery probability is

$$\Pr[g(V) = Y] = \sum_o \sum_{y \in \mathcal{F}(o)} \pi(y) q_o(y) \tag{40}$$

$$\le \sum_o \max_{y \in \mathcal{F}(o)} \pi(y), \tag{41}$$

because $\sum_y q_o(y) = 1$. Equality is attained by assigning all output probability to a posterior maximizer in each fiber. □

Remark 5.3 (General jointly distributed view). For arbitrary jointly distributed finite random variables (Y, V), even when V is not a deterministic function of Y, the same decision argument gives

$$P_{\text{guess}}(Y \mid V) = \sum_v \Pr[V = v] \max_y \Pr[Y = y \mid V = v] = \sum_v \max_y \Pr[Y = y, V = v]. \tag{42}$$

Theorem 5.2 is the deterministic specialization: when $V = O(Y)$, each posterior is supported on a fiber and Equation (42) reduces to Equation (38). Correlated non-deterministic side information updates the posterior through the joint law and need not refine a deterministic fiber partition.

Corollary 5.4 (Uniform finite domain)**.** *Suppose Y is uniform on $\mathcal{S}$, and write*

$$N = |\mathcal{S}|, \qquad m = |O(\mathcal{S})|, \qquad k_o = |\mathcal{F}(o)|. \tag{43}$$

Then

$$P_{\text{guess}}(Y \mid V) = \frac{m}{N}. \tag{44}$$

Conditioned on a particular attainable observation o,

$$\max_y \Pr[Y = y \mid V = o] = \frac{1}{k_o}. \tag{45}$$

Proof. The posterior is uniform on each fiber, so every fiber contributes exactly $1/N$ to Equation (38). Summing over the m attainable observations gives Equation (44). The pointwise result follows because all k_o members have posterior probability $1/k_o$. □

Equation (44) is an exact optimum, not an upper bound on one selected attack. It describes an ideal decoder that knows the declared finite domain, prior, complete observation map, and every fiber.

5.2 Conditional entropy and mutual information

All logarithms in this section have base two. The conditional Shannon entropy is

$$H(Y \mid V) = \sum_o p_o H(\pi_o) \tag{46}$$

$$= -\sum_{y \in \mathcal{S}_\pi} \pi(y) \log_2 \left(\frac{\pi(y)}{p_{O(y)}} \right). \tag{47}$$

Because $V = O(Y)$ is deterministic in this section, $H(V \mid Y) = 0$ and therefore

$$I(Y; V) = H(V) - H(V \mid Y) = H(V). \tag{48}$$

Proposition 5.5 (Deterministic views and perfect secrecy)**.** *Let Y be finite valued with positive-probability support $\mathcal{S}_\pi$, and let $V = O(Y)$ be deterministic. Then the following are equivalent:*

1. *$I(Y; V) = 0$;*
2. *Y and V are independent;*
3. *V is almost surely constant;*
4. *O is constant on $\mathcal{S}_\pi$.*

Thus a deterministic view provides Shannon perfect secrecy about the complete input under the declared prior if and only if it is constant on that prior's support. If this must hold for every prior supported on a domain $\mathcal{S}$, then O must be constant on all of $\mathcal{S}$.

Proof. Equation (48) gives $I(Y; V) = H(V)$. A finite-valued random variable has zero entropy exactly when its distribution is concentrated at one value, which is equivalent to O being constant on $\mathcal{S}_\pi$. Zero mutual information is equivalent to independence. Finally, if O is nonconstant on $\mathcal{S}$, a prior assigning positive probability to two inputs with different outputs gives $H(V) > 0$; the converse is immediate. □

Proposition 5.6 (Uniform-prior Shannon information)**.** *Under the uniform prior of Corollary 5.4,*

$$H(Y \mid V) = \frac{1}{N} \sum_{o \in O(\mathcal{S})} k_o \log_2 k_o = \mathbb{E}[\log_2 |\mathcal{F}(V)|], \tag{49}$$

$$I(Y;V) = H(Y) - H(Y \mid V) \tag{50}$$

$$= \log_2 N - \frac{1}{N} \sum_o k_o \log_2 k_o \tag{51}$$

$$= H(V) = -\sum_o \frac{k_o}{N} \log_2 \left(\frac{k_o}{N} \right). \tag{52}$$

Proof. Given $V = o$, the posterior is uniform on $\mathcal{F}(o)$ and has entropy $\log_2 k_o$. The observation probability is $p_o = k_o/N$, which proves Equation (49). The mutual-information identities follow from $H(Y) = \log_2 N$ and $H(V \mid Y) = 0$. □

For a particular observation o, define the pointwise conditional min-entropy

$$H_\infty(Y \mid V = o) = -\log_2 \max_y \Pr[Y = y \mid V = o]. \tag{53}$$

The operational average conditional min-entropy [3] is

$$\widetilde{H}_\infty(Y \mid V) = -\log_2 \sum_o p_o 2^{-H_\infty(Y|V=o)} \tag{54}$$

$$= -\log_2 P_{\text{guess}}(Y \mid V). \tag{55}$$

Under a uniform prior,

$$H_\infty(Y \mid V = o) = \log_2 k_o, \tag{56}$$

$$\widetilde{H}_\infty(Y \mid V) = \log_2 \frac{N}{m}. \tag{57}$$

The tilde is material: average conditional min-entropy is the negative logarithm of the average optimal guessing probability, not the arithmetic mean of the pointwise min-entropies.

5.3 Candidate-set distribution

Define the candidate-set-size random variable

$$K = |\mathcal{F}(V)|. \tag{58}$$

For a general prior,

$$\Pr[K = j] = \sum_{\substack{o \in O(\mathcal{S}_\pi) \\ |\mathcal{F}(o)| = j}} p_o, \tag{59}$$

$$\mathbb{E}[K] = \sum_o p_o |\mathcal{F}(o)|. \tag{60}$$

Proposition 5.7 (Uniform-prior fiber-size distribution)**.** *Under a uniform prior, let*

$$n_j = |\{o \in O(\mathcal{S}) : k_o = j\}|. \tag{61}$$

Then

$$\sum_{j\geq 1} n_j = m, \tag{62}$$

$$\sum_{j\geq 1} j n_j = N, \tag{63}$$

$$\Pr[K = j] = \frac{j n_j}{N}, \tag{64}$$

$$\mathbb{E}[K] = \frac{1}{N}\sum_o k_o^2 = \frac{1}{N}\sum_{j\geq 1} j^2 n_j. \tag{65}$$

The probability that a random input lies in a singleton fiber is n_1/N; the non-singleton probability is $1 - n_1/N$.

Proof. There are n_j fibers of size j, containing jn_j inputs in total. Under a uniform input prior, those inputs have probability jn_j/N. The remaining identities follow by summation. □

The distribution in Equation (64) is size-biased because a random input is more likely to lie in a large fiber. If an attainable observation were instead selected uniformly, the mean fiber size would be N/m. These averages answer different questions and must not be interchanged.

Corollary 5.8 (Constant-view prior-only baseline)**.** *Suppose the observer receives neither C nor R, and every other in-scope released field is constant on the positive-probability support. Equivalently, for some fixed value $v_\varnothing$,*

$$V_\varnothing = O_\varnothing(Y) = v_\varnothing \qquad \textit{almost surely.} \tag{66}$$

Then

$$\mathcal{F}_\varnothing(v_\varnothing) = \mathcal{S}_\pi, \tag{67}$$

$$P_{\text{guess}}(Y \mid V_\varnothing) = \max_{y\in\mathcal{S}_\pi} \pi(y), \tag{68}$$

$$\widetilde{H}_\infty(Y \mid V_\varnothing) = -\log_2 \max_{y\in\mathcal{S}_\pi} \pi(y), \tag{69}$$

$$H(Y \mid V_\varnothing) = H(Y), \tag{70}$$

$$I(Y; V_\varnothing) = 0, \tag{71}$$

$$K = |\mathcal{S}_\pi| \qquad \textit{almost surely.} \tag{72}$$

If Y is uniform on a domain $\mathcal{S}$ of size N, then

$$P_{\text{guess}}(Y \mid V_\varnothing) = \frac{1}{N}, \qquad \widetilde{H}_\infty(Y \mid V_\varnothing) = H(Y \mid V_\varnothing) = \log_2 N, \qquad I(Y; V_\varnothing) = 0, \qquad \mathbb{E}[K] = N. \tag{73}$$

Proof. The constant map has one attainable observation and its sole fiber is $\mathcal{S}_\pi$. Substitution into Theorem 5.2, Equation (55), and Equation (60) gives the guessing, min-entropy, and candidate-set results. Conditioning on a constant random variable leaves the distribution of Y unchanged, so $H(Y \mid V_\varnothing) = H(Y)$ and $I(Y; V_\varnothing) = 0$. □

The premise concerns the complete cumulative view, not merely the absence of the two numerical arrays. If the observer receives data-dependent metadata $M(Y)$, then the relevant view is $V_M = M(Y)$ and its fibers must be analyzed separately. Correlated side information Z produces the view $(V_\varnothing, Z)$, which is equivalent to Z and is governed by Remark 5.3.

5.4 Other loss functions and sensitive attributes

The same partition supports Bayes-optimal analysis for other declared targets. For an action space $\mathcal{A}$ and loss $\ell(y, a)$,

$$R^*_\ell(Y \mid V) = \inf_g \mathbb{E}[\ell(Y, g(V))] = \sum_o \inf_{a\in\mathcal{A}} \sum_{y\in\mathcal{F}(o)} \pi(y)\ell(y, a). \tag{74}$$

Squared error, approximate matching, and classification require different losses and can therefore give conclusions very different from exact recovery.

Proposition 5.9 (Bayes-optimal squared-Frobenius reconstruction)**.** *Suppose Y is matrix valued with finite second moment, let possibly randomized decoder actions range over the ambient Euclidean matrix space, and score them by squared Frobenius loss. Then the conditional mean*

$$g^*(V) = \mathbb{E}[Y \mid V] \tag{75}$$

is Bayes optimal, and its minimum risk is

$$R^*_{\mathrm{SE}}(Y \mid V) = \mathbb{E}\Big[\|Y - \mathbb{E}[Y \mid V]\|_F^2\Big] \tag{76}$$

$$= \mathbb{E}[\operatorname{tr} \operatorname{Cov}(\operatorname{vec}(Y) \mid V)]\,. \tag{77}$$

For a finite deterministic view, define

$$\mu_o = \sum_{y \in \mathcal{F}(o)} \pi_o(y) y. \tag{78}$$

Then

$$R^*_{\mathrm{SE}}(Y \mid V) = \sum_o p_o \sum_{y \in \mathcal{F}(o)} \pi_o(y) \|y - \mu_o\|_F^2. \tag{79}$$

Under a uniform prior this becomes

$$R^*_{\mathrm{SE}}(Y \mid V) = \frac{1}{N} \sum_o \sum_{y \in \mathcal{F}(o)} \|y - \mu_o\|_F^2, \qquad \mu_o = \frac{1}{k_o} \sum_{y \in \mathcal{F}(o)} y. \tag{80}$$

Proof. Let $\mu_V = \mathbb{E}[Y \mid V]$. Conditional on V, the standard orthogonal decomposition gives

$$\mathbb{E}[\|Y - a\|_F^2 \mid V] = \mathbb{E}[\|Y - \mu_V\|_F^2 \mid V] + \|a - \mu_V\|_F^2. \tag{81}$$

The second term is minimized at $a = \mu_V$. Randomization averages conditional risks and cannot improve this minimum. Averaging over V proves the first identity; the second is the coordinate form of the conditional variance. Expanding the conditional expectations gives Equations (79) and (80). □

If decoder outputs are required to lie in a restricted action set, the conditional mean need not be admissible. Under squared loss, an optimal restricted action is a point in that set nearest to the conditional mean.

5.5 Isotropic Gaussian prior

Theorem 5.10 (Gaussian posterior and reconstruction bounds)**.** *Let*

$$X \sim \mathcal{N}(\mu, \sigma^2 I_n), \qquad \sigma > 0, \qquad V = AX, \tag{82}$$

where $A \in \mathbb{R}^{m \times n}$ is fixed. Write

$$\Pi_A = A^+ A, \qquad N_A = I_n - \Pi_A, \qquad q = n - \operatorname{rank}(A). \tag{83}$$

For every v in the support of V, the Gaussian regular conditional distribution is

$$X \mid V = v \sim \mathcal{N}\big(\mu + A^+(v - A\mu), \sigma^2 N_A\big)\,. \tag{84}$$

The posterior covariance has rank q, and the Bayes-optimal total squared-error risk is

$$R^*_{\mathrm{SE}}(X \mid V) = \sigma^2 q. \tag{85}$$

Let $F_{\chi^2_q}$ denote the cumulative distribution function of a chi-squared random variable with q degrees of freedom, with χ^2_0 interpreted as a point mass at zero. For every $\tau \geq 0$ and every possibly randomized decoder satisfying the Markov condition $X \to V \to \widehat{X}$,

$$\Pr\Big[\|X - \widehat{X}\|_2 \leq \tau \;\Big|\; V = v\Big] \leq F_{\chi^2_q}\left(\frac{\tau^2}{\sigma^2}\right). \tag{86}$$

Equality is attained by the posterior-mean decoder

$$\widehat{X}^*(v) = \mu + A^+(v - A\mu). \tag{87}$$

Consequently, if $q > 0$, every decoder has zero exact point-recovery probability.

If, under the same Markov condition, a decoder returns a possibly randomized list $\widehat{\mathcal{L}}$ of at most ℓ reconstructions, then

$$\Pr\left[\inf_{\widehat{x}\in\widehat{\mathcal{L}}}\|X - \widehat{x}\|_2 \leq \tau \;\middle|\; V = v\right] \leq \min\left\{1, \ell F_{\chi^2_q}\left(\frac{\tau^2}{\sigma^2}\right)\right\}. \tag{88}$$

Proof. Let $Z = X - \mu$. The matrices Π_A and N_A are the orthogonal projectors onto $\ker(A)^\perp$ and $\ker(A)$, respectively. The Gaussian vectors $\Pi_A Z$ and $N_A Z$ have zero cross-covariance,

$$\mathrm{Cov}(\Pi_A Z, N_A Z) = \sigma^2 \Pi_A N_A = 0, \tag{89}$$

and are therefore independent. The observation determines the first component because

$$\Pi_A Z = A^+(V - A\mu), \tag{90}$$

while the second retains distribution $\mathcal{N}(0, \sigma^2 N_A)$. This proves Equation (84). Its covariance trace is $\sigma^2 \operatorname{tr}(N_A) = \sigma^2 q$, so Proposition 5.9 proves Equation (85).

Conditioned on $V = v$, the posterior is an isotropic Gaussian on a q-dimensional affine subspace. Orthogonally projecting any proposed reconstruction onto that affine subspace cannot increase its distance from X. Within the subspace, a radius-τ Gaussian ball has maximum probability when centered at the posterior mean [6]. At that center,

$$\frac{\|X - \widehat{X}^*(v)\|_2^2}{\sigma^2} \sim \chi^2_q, \tag{91}$$

which proves Equation (86). A randomized decoder averages fixed-center probabilities and therefore cannot improve the bound. When $q > 0$, the posterior is non-atomic on its support, giving zero exact point-recovery probability. Finally, apply the union bound to the at most ℓ posterior balls generated by each realized list and average over decoder randomness to obtain Equation (88). □

Corollary 5.11 (Gaussian coefficient and residual recovery)**.** *Suppose*

$$\mathrm{vec}(Y) \sim \mathcal{N}\big(\mathrm{vec}(M), \sigma^2 I_{LD}\big). \tag{92}$$

For a fixed basis B of rank r and every c and ρ in the supports of C and R, respectively,

$$\mathrm{vec}(Y) \mid C = c \sim \mathcal{N}\big(\mathrm{vec}(Bc + QM), \sigma^2(I_D \otimes Q)\big), \tag{93}$$

$$\mathrm{vec}(Y) \mid R = \rho \sim \mathcal{N}\big(\mathrm{vec}(\rho + PM), \sigma^2(I_D \otimes P)\big). \tag{94}$$

Therefore the Bayes-optimal total squared-error risks are

$$R^*_{\mathrm{SE}}(Y \mid C) = \sigma^2(L - r)D, \tag{95}$$

$$R^*_{\mathrm{SE}}(Y \mid R) = \sigma^2 r D. \tag{96}$$

For every $\tau \geq 0$, the optimal single-reconstruction probabilities are exactly

$$\sup_G \Pr[\|Y - \widehat{Y}\|_F \leq \tau \mid C = c] = F_{\chi^2_{(L-r)D}}\left(\frac{\tau^2}{\sigma^2}\right), \tag{97}$$

$$\sup_G \Pr[\|Y - \widehat{Y}\|_F \leq \tau \mid R = \rho] = F_{\chi^2_{rD}}\left(\frac{\tau^2}{\sigma^2}\right). \tag{98}$$

Under this prior, C and R are independent.

Proof. For coefficients, apply Theorem 5.10 with $A = I_D \otimes B^+$. Its rank is rD, and

$$I_{LD} - A^+A = I_D \otimes Q. \tag{99}$$

For residuals, take $A = I_D \otimes Q$, whose rank is $(L-r)D$ and whose null-space projector is $I_D \otimes P$. The displayed posterior, squared-error, and small-ball formulas follow. Finally,

$$\mathrm{Cov}(\mathrm{vec}(C), \mathrm{vec}(R)) = \sigma^2(I_D \otimes B^+Q) = 0. \tag{100}$$

The pair is jointly Gaussian, so zero cross-covariance implies independence. □

The independence statement is specific to the fixed-basis isotropic Gaussian prior. It is not a distribution-free secrecy claim and may fail under a non-isotropic or non-Gaussian population. The general theorem also covers repeated fixed linear releases by taking A to be the stacked cumulative operator from Theorem 4.2.

In particular, input non-identifiability does not by itself guarantee protection of an attribute. Let $A = h(Y)$ take values in a finite set $\mathcal{A}$. Its optimal recovery probability is

$$P_{\mathrm{guess}}(A \mid V) = \sum_o \max_{a \in \mathcal{A}} \sum_{\substack{y \in \mathcal{F}(o) \\ h(y)=a}} \pi(y). \tag{101}$$

Under a uniform prior this becomes

$$P_{\mathrm{guess}}(A \mid V) = \frac{1}{N} \sum_o \max_{a \in \mathcal{A}} \left|\{y \in \mathcal{F}(o) : h(y) = a\}\right|. \tag{102}$$

A large input fiber can therefore reveal an attribute perfectly when that attribute is constant throughout the fiber.

Every quantity in this section is conditional on its declared finite domain, prior, quantization, and observer view. Section 14 states the boundaries on broader interpretations.

6 Policy-Selected Disclosure and Split Custody

The component model supports more than a binary choice between releasing and withholding a token. A policy may select among several complete observer views, and complementary channels may be administered by different parties. The resulting guarantee must be stated for each selected view and each possible coalition.

Selected disclosure controls which observation map becomes available. It does not, by itself, establish that every selected view has low leakage. That conclusion remains conditional on the declared domain, prior, target, metric, side information, and cumulative release history.

Definition 6.1 (Policy-selected disclosure order)**.** Let $\mathcal{D}$ be a set of disclosure labels. For every $d \in \mathcal{D}$, let

$$O_d : \mathcal{S} \longrightarrow \mathcal{V}_d, \qquad V_d = O_d(Y), \tag{103}$$

be the complete deterministic view released under policy choice d. The view includes every numerical channel, metadata field, selection label, certificate, and earlier release visible to the observer. If the policy selector $\delta : \mathcal{S} \to \mathcal{D}$ depends on Y, the released view is

$$\left(\delta(Y), O_{\delta(Y)}(Y)\right), \tag{104}$$

not merely $O_d(Y)$ conditional on the realized label.

Write $d \preceq e$ when there exists a deterministic map $h_{d \leftarrow e}$ such that

$$O_d = h_{d \leftarrow e} \circ O_e. \tag{105}$$

Thus an observer receiving V_e can compute everything available in V_d. After informationally equivalent views are identified, $\preceq$ is a partial order. We use the same notation for view maps when one factors deterministically through the other.

With every preexisting observable field, preprocessing choice, and release label held fixed, selecting additional channels gives a sufficient refinement relation. The coefficient-only and residual-only views, however, are generally incomparable: neither is normally a deterministic function of the other. Both are refined by the joint view,

$$V_\varnothing \preceq V_C \preceq V_{C,R}, \qquad V_\varnothing \preceq V_R \preceq V_{C,R}. \tag{106}$$

Selected disclosure is therefore a partial order rather than a universal one-dimensional security dial.

Theorem 6.2 (Monotonicity under disclosure refinement)**.** *Let $d \preceq e$, let Y have a declared finite prior on $\mathcal{S}$, and let $U = \phi(Y)$ be any finite-valued target. Then, for every $y \in \mathcal{S}$,*

$$\mathcal{F}_e(O_e(y)) \subseteq \mathcal{F}_d(O_d(y)). \tag{107}$$

Using Definition 5.1 with U in place of Y,

$$P_{\text{guess}}(U \mid V_d) \le P_{\text{guess}}(U \mid V_e), \tag{108}$$
$$H(U \mid V_d) \ge H(U \mid V_e), \tag{109}$$
$$I(U; V_d) \le I(U; V_e), \tag{110}$$
$$\widetilde{H}_\infty(U \mid V_d) \ge \widetilde{H}_\infty(U \mid V_e). \tag{111}$$

For every declared loss ℓ and common action space, the Bayes-optimal risks satisfy

$$R^*_\ell(U \mid V_e) \le R^*_\ell(U \mid V_d). \tag{112}$$

For the complete-input target $U = Y$, the incremental mutual information is exactly

$$I(Y; V_e) - I(Y; V_d) = H(V_e \mid V_d). \tag{113}$$

Proof. Equation (105) implies that equality of two e-views forces equality of their d-views, proving Equation (107). An observer of V_e can reproduce any decoder or action based on V_d by first computing $h_{d \leftarrow e}(V_e)$. Additional information therefore cannot reduce optimal guessing probability or increase optimal Bayes risk. Equation (111) follows from Equation (55). Because V_d is a deterministic function of V_e, data processing and the chain rule prove Equations (109) and (110). When $U = Y$, both views are deterministic functions of Y, so

$$I(Y; V_e \mid V_d) = H(V_e \mid V_d), \tag{114}$$

which proves Equation (113). □

The theorem supplies the precise sense in which a policy controls disclosure: moving upward in the order can only shrink or preserve ambiguity and cannot worsen an optimal observer's inference performance. It does not rank incomparable views. Coefficients and residuals must be compared using explicit recovery, entropy, attribute, or task-utility measurements.

6.1 Coalition views

Definition 6.3 (Split-custody channel assignment)**.** Let $\mathcal{P}$ be a finite party set, let $O_0(Y)$ be the complete public transcript, and let $\{O_j(Y) : j \in \mathcal{J}\}$ be separately administered channels. Assign each party $p \in \mathcal{P}$ a channel set $J_p \subseteq \mathcal{J}$. For a coalition $\Gamma \subseteq \mathcal{P}$, define

$$J_\Gamma = \bigcup_{p \in \Gamma} J_p, \tag{115}$$
$$O_\Gamma(Y) = \Big(O_0(Y), \big(O_j(Y)\big)_{j \in J_\Gamma}\Big). \tag{116}$$

The map O_Γ is the coalition's complete cumulative view. This deterministic channel model applies directly to component and metadata observations. If encryption, commitments, confidential execution, or a protocol uses fresh randomness, the resulting coalition observation must instead be analyzed through its joint law with Y as in Remark 5.3, including every observable ciphertext and transcript field.

Theorem 6.4 (Coalition fibers and functional access)**.** *For each public or assigned channel, define the input-indexed fiber*

$$\mathcal{F}_j(y) = \{x \in \mathcal{S} : O_j(x) = O_j(y)\}, \tag{117}$$

and define $\mathcal{F}_\Gamma(y)$ analogously for O_Γ. Then, for every $y \in \mathcal{S}$,

$$\mathcal{F}_\Gamma(y) = \mathcal{F}_0(y) \cap \bigcap_{j \in J_\Gamma} \mathcal{F}_j(y), \tag{118}$$

where each fiber fixes the corresponding component of the realized view. If $\Gamma \subseteq \Lambda$, then

$$O_\Gamma \preceq O_\Lambda, \qquad \mathcal{F}_\Lambda(y) \subseteq \mathcal{F}_\Gamma(y). \tag{119}$$

Let $U = \phi(Y)$ be the target a designated coalition is intended to recover. The following are equivalent:

1. *there exists a decoder g_Γ such that $g_\Gamma(O_\Gamma(y)) = \phi(y)$ for every $y \in \mathcal{S}$, equivalently $\phi = g_\Gamma \circ O_\Gamma$; and*
2. *ϕ is constant on every attainable coalition fiber.*

For full-input recovery, ϕ is the identity map and the criterion reduces to every coalition fiber being a singleton.

Proof. Equality of the coalition tuple is equivalent to equality of every public and assigned channel, which proves Equation (118). A larger coalition has a superset of the channels, proving Equation (119). If a decoder exists, two inputs in the same coalition fiber must have the same target value. Conversely, if ϕ is constant on every fiber, define $g_\Gamma(v)$ as that common value for each attainable coalition observation v. This decoder is well defined and proves the factorization. □

The decoder-existence statement is set-theoretic. It does not establish computational efficiency. On an infinite measurable domain, an operational probabilistic claim also requires the relevant maps and decoder to be measurable.

Coalitions designated to receive reconstructing channel sets form a monotone functional access structure: once a coalition can recover the target, every larger coalition can also recover it. Failure of factorization for a coalition proves only that universal exact recovery is impossible. A quantitative access policy must additionally bound that coalition's guessing probability, mutual information, Bayes risk, attribute inference, or other declared measure.

Corollary 6.5 (Linear coalition criterion)**.** *Suppose the public and assigned channels are fixed linear maps acting columnwise, and let A_Γ be their vertical stack. On the unrestricted fitting domain,*

$$\dim \mathcal{F}_\Gamma = D\,(L - \operatorname{rank}(A_\Gamma))\,. \tag{120}$$

The coalition reconstructs every fitting-domain input exactly if and only if $\ker(A_\Gamma) = \{0\}$. On a restricted domain $\mathcal{S}$, exact reconstruction holds if and only if

$$(\mathcal{S} - \mathcal{S}) \cap \ker(A_\Gamma)^D = \{0\}. \tag{121}$$

Proof. Apply Theorem 4.2 to the stacked coalition operator, then apply its unrestricted and restricted injectivity criteria. □

Corollary 6.6 (Coefficient-residual staged custody)**.** *Fix a basis B of rank r and let $Q = I - BB^+$. Suppose a receiver is given the exact pre-serialization component $C = B^+Y$, a custodian holds the exact plaintext residual $R = QY$, and B is available to the reconstructing coalition. Before the residual channel is released, the receiver's unrestricted fitting-domain fiber has dimension $(L-r)D$. Because the custodian observes plaintext R, its residual-only fiber has dimension rD. Once policy permits the exact residual to be joined with the receiver's exact coefficient view, the joint fiber is a singleton and*

$$Y = BC + R. \tag{122}$$

Canonical source-domain inversion may additionally consume inverse-preprocessing fields. Whether their absence prevents source recovery is governed by the source-domain fiber, and some preprocessing may remain noninvertible even when every recorded field is present.

An illustrative three-role arrangement gives exact coefficients and task metadata to the receiver, places the exact residual under a custodian's release control, and leaves fields used by the canonical source-domain inverse with the originator. After coefficient-residual recombination, those fields may stage the intended source-domain inversion. Whether withholding them actually prevents source recovery is a source-domain fiber question. The arrangement is not automatically a secure three-of-three scheme: every proper downstream coalition must still be analyzed, the inverse metadata may itself leak, and the originator normally knew the source before decomposition. The originator's approval is therefore a downstream release gate, not an independent secret share protecting the source from the originator. If the custodian controls only a ciphertext, wrapped key, credential, or execution rule rather than plaintext R, its view must be modeled from those actual holdings instead of the residual fiber. If a smart contract mediates that role, its observable state and execution transcript belong to O_0; confidential custody requires a separate encryption, threshold-key, confidential-execution, or trusted off-chain mechanism.

7 Approximate and Worst-Case Reconstruction Bounds

Non-injectivity implies a quantitative lower bound for every deterministic or randomized decoder.

Theorem 7.1 (Two-point reconstruction lower bound)**.** *Let $\mathcal{S}$ lie in a reconstruction space with metric d, let $O : \mathcal{S} \to \mathcal{O}$ be any observation map, and suppose $Y_0 \neq Y_1$ satisfy $O(Y_0) = O(Y_1) = o$. Let G be any possibly randomized decoder with output in that metric space, and let $\widehat{Y} \sim G(\cdot \mid o)$. Then*

$$\max\left\{\mathbb{E}_G[d(\widehat{Y}, Y_0)], \mathbb{E}_G[d(\widehat{Y}, Y_1)]\right\} \geq \frac{1}{2} d(Y_0, Y_1). \tag{123}$$

For a deterministic decoder, the expectations reduce to the two pointwise errors.

Proof. Because the two observations agree, the conditional law of $\widehat{Y}$ is the same under Y_0 and Y_1. For every realized output, the triangle inequality gives

$$d(Y_0, Y_1) \leq d(Y_0, \widehat{Y}) + d(\widehat{Y}, Y_1). \tag{124}$$

Integrating with respect to the common output law shows that the two expected errors sum to at least $d(Y_0, Y_1)$. At least one is therefore at least half that distance. □

Writing

$$\mathcal{F}_O(o) = \{Y \in \mathcal{S} : O(Y) = o\}, \tag{125}$$

the same argument gives the randomized minimax form

$$\inf_G \sup_{Y \in \mathcal{S}} \mathbb{E}_{\widehat{Y} \sim G(\cdot \mid O(Y))}[d(\widehat{Y}, Y)] \geq \frac{1}{2} \sup_o \operatorname{diam} \mathcal{F}_O(o). \tag{126}$$

For unrestricted real-valued inputs, the relevant coefficient and residual fibers are unbounded whenever Corollary 3.2 says the map is non-injective. Their worst-case Frobenius reconstruction error is therefore unbounded.

Theorem 7.2 (Separated-fiber list-reconstruction bound)**.** *Let $(\mathcal{X}, d)$ be a metric reconstruction space, let $O : \mathcal{S} \to \mathcal{O}$, and suppose that $Y_1, \ldots, Y_K \in \mathcal{S}$ satisfy*

$$O(Y_1) = \cdots = O(Y_K) = o, \qquad d(Y_i, Y_j) \geq 2\tau \quad (i \neq j), \tag{127}$$

for some $\tau > 0$. Let $q \in \mathbb{Z}_{\geq 0}$. On observing o, let a possibly randomized decoder return a list $\widehat{\mathcal{L}}_o \sim G(\cdot \mid o)$ containing at most q points of $\mathcal{X}$. Its internal randomness is independent of the source conditional on the observation. Define

$$s_i = \Pr_G\left[\inf_{\widehat{Y} \in \widehat{\mathcal{L}}_o} d(\widehat{Y}, Y_i) < \tau\right], \tag{128}$$

with the infimum over an empty list defined as $+\infty$. Then

$$\sum_{i=1}^{K} s_i \leq \min\{K, q\}, \qquad \min_{1 \leq i \leq K} s_i \leq \min\left\{1, \frac{q}{K}\right\}. \tag{129}$$

The worst-case bound is sharp:

$$\sup_G \min_{1\le i\le K} s_i = \min\Big\{1, \frac{q}{K}\Big\}. \tag{130}$$

If Y is random, $V = O(Y)$, and conditional on $V = o$ the posterior support is exactly $\{Y_1, \dots, Y_K\}$ with probabilities $p_{(1)} \ge \cdots \ge p_{(K)}$, then

$$\sup_G \Pr\left[\inf_{\widehat{Y}\in\widehat{\mathcal{L}}} d(Y,\widehat{Y}) < \tau \;\middle|\; V = o\right] = \sum_{j=1}^{\min\{K,q\}} p_{(j)}. \tag{131}$$

In particular, under the uniform conditional prior the optimum is $\min\{1, q/K\}$.

Proof. For a realized list $\mathcal{L}$, define

$$I_i(\mathcal{L}) = \mathbf{1}\left\{\inf_{\widehat{Y}\in\mathcal{L}} d(\widehat{Y}, Y_i) < \tau\right\}. \tag{132}$$

No single list element can satisfy the displayed inequality for two different Y_i and Y_j: otherwise the triangle inequality would give $d(Y_i, Y_j) < 2\tau$, contradicting Equation (127). Consequently,

$$\sum_{i=1}^{K} I_i(\mathcal{L}) \le \min\{K, |\mathcal{L}|\} \le \min\{K, q\}. \tag{133}$$

The decoder has the same output law under every Y_i because all Y_i produce the same observation. Taking expectation over that common law proves Equation (129).

For $q < K$, a decoder that selects a uniformly random q-element subset of $\{Y_1, \dots, Y_K\}$ succeeds on every Y_i with probability q/K. For $q \ge K$, it can return all K points. This proves Equation (130). For the posterior statement, every realized list covers at most q posterior support points. Its covered posterior mass is therefore at most the sum of the q largest posterior probabilities, and returning those points attains the bound. Randomization only averages covered masses and cannot improve it. □

The strict error event $d < \tau$ permits the weak separation $d(Y_i, Y_j) \ge 2\tau$. A closed-ball success event $d \le \tau$ instead requires strict separation $d(Y_i, Y_j) > 2\tau$.

Corollary 7.3 (Squared-Frobenius diameter bound)**.** *In a Euclidean matrix space, any two inputs Y_0, Y_1 with the same observation satisfy, for every possibly randomized decoder,*

$$\max_{i\in\{0,1\}} \mathbb{E}_G\Big[\|\widehat{Y} - Y_i\|_F^2\Big] \ge \frac{1}{4}\|Y_0 - Y_1\|_F^2. \tag{134}$$

Consequently,

$$\inf_G \sup_{Y\in\mathcal{S}} \mathbb{E}_G\Big[\|\widehat{Y} - Y\|_F^2\Big] \ge \frac{1}{4}\sup_o \operatorname{diam}(\mathcal{F}_O(o))^2. \tag{135}$$

Proof. Because the conditional law of $\widehat{Y}$ is the same for the two indistinguishable inputs,

$$\frac{1}{2}\mathbb{E}_G\Big[\|\widehat{Y} - Y_0\|_F^2 + \|\widehat{Y} - Y_1\|_F^2\Big] = \mathbb{E}_G\left[\left\|\widehat{Y} - \frac{Y_0 + Y_1}{2}\right\|_F^2\right] + \frac{1}{4}\|Y_0 - Y_1\|_F^2 \tag{136}$$

$$\ge \frac{1}{4}\|Y_0 - Y_1\|_F^2. \tag{137}$$

At least one of the two expected losses is no smaller than their average. Taking pairs arbitrarily close to each fiber diameter and then the supremum over observations proves Equation (135). □

Under the discrete metric, Theorem 7.1 gives the prior-free two-candidate error statement, while Theorem 7.2 gives the exact uniform-posterior extension to separated K-candidate fibers and list reconstruction. Theorem 5.2 remains the general prior-conditional exact-recovery result.

8 Effects of Metadata and Auxiliary Fields

8.1 No-component and metadata-only observations

The absence of both numerical arrays does not by itself define an empty observation. Let

$$V_{\varnothing} = v_{\varnothing} \tag{138}$$

denote a complete cumulative view that is constant on the declared prior support, and let

$$V_M = M(Y) \tag{139}$$

denote the complete metadata and observable transcript remaining when neither C nor R is released. Corollary 5.8 applies to $V_{\varnothing}$. Theorem 5.2 applies to V_M after its fields and equality rule are frozen. Under a uniform prior, if M has m_M attainable values on a domain of size N, then

$$P_{\text{guess}}(Y \mid V_M) = \frac{m_M}{N}, \qquad \widetilde{H}_{\infty}(Y \mid V_M) = \log_2 \frac{N}{m_M}. \tag{140}$$

Without restrictions on M, its leakage can range from zero to complete identification. Fixed public configuration contributes no incremental information after conditioning on that configuration. A data-selected basis, boundary, normalization value, magnitude, digest, output length, error, or release decision is part of $M(Y)$ and must not be treated as fixed merely because it is called metadata.

The finite-domain evidence below also evaluates one configuration-scoped logical decoded metadata view. With one fixed five-row, one-column, degree-one segment and preprocessing disabled, all decoded structural fields are constant and the source-dependent portion is

$$V_M^*(Y) = \left(D_P(Y), S_R(Y)\right), \tag{141}$$

where D_P is the deterministic digest of the integer-scaled fitted prediction and S_R is the IEEE-754 binary32 representation of the decoded residual-magnitude scalar. The replay uses exact byte equality for D_P and S_R and reproduces the source equations and one-column operation shapes specified in Appendix A. It omits both coefficient and residual arrays. It also excludes the file-wide residual-quantization scale, token length, timing, errors, later releases, and any other field not returned in the declared logical decoded metadata view. The recorded MLX 0.31.2 runtime matches the earlier controlled witness environment but falls outside the referenced source tag's declared requirement of MLX $\geq 0.30.6, < 0.31$. These exclusions and the runtime mismatch prevent Table 6 from being read as a complete serialized-package or release-conformance result.

8.2 Released residual magnitude

To avoid dependence on any particular encoding or normalization factor, grant the observer the exact Frobenius magnitude

$$a = \|R\|_F. \tag{142}$$

This is at least as informative as any disclosed scalar that is a known deterministic function of $\|R\|_F$ alone: two residuals with the same exact norm necessarily have the same value of every such field.

Suppose B, C, and a are known, and put

$$n = (L - r)D. \tag{143}$$

The compatible fitting-domain inputs are

$$\mathcal{F}_{C,a} = BC + \left\{ Z \in \mathbb{R}^{L \times D} : B^T Z = 0,\ \|Z\|_F = a \right\}. \tag{144}$$

Consequently:

- if $a = 0$, then $R = 0$ and $Y = BC$ is unique when B and inverse preprocessing are known;
- if $a > 0$ and $n = 1$, the compatible set consists of two points;

- if $a > 0$ and $n \geq 2$, the compatible residuals form an $(n-1)$-dimensional sphere in $\ker(B^T)^D$;
- if $n = 0$, only $a = 0$ is attainable.

Thus exact residual magnitude removes the radial degree of freedom but generally leaves the angular ambiguity. Since this conclusion grants more norm information than a coarser norm-only field, it also applies to any such field. Beside the exact pre-serialization residual, every norm-only field that is a known function of R is redundant. A magnitude computed before quantization may add information beside a quantized residual.

8.3 Conditional deterministic prediction field

Suppose a protocol conditionally releases a deterministic field derived from the fitted prediction:

$$d = \Gamma(\mathcal{Q}_P(BC)), \tag{145}$$

where $\mathcal{Q}_P$ and Γ are deterministic maps. No cryptographic property is assumed. Their exact transforms and representations are intentionally outside this edition.

When B and C are already known, d is determined by the known prediction and supplies no residual information. With a residual observation, the joint field may narrow the compatible inputs to

$$\{R + BA : \Gamma(\mathcal{Q}_P(BA)) = d\} \cap \mathcal{S}. \tag{146}$$

It may also support offline verification of low-entropy guesses. Neither property implies that the field uniquely identifies the fitted component.

A conditional collision result requires only local stability, not disclosure of the transform. Assume there is a radius $\rho_{\mathcal{Q}_P}(BC) > 0$ such that

$$\|U - BC\|_\infty < \rho_{\mathcal{Q}_P}(BC) \quad \Longrightarrow \quad \mathcal{Q}_P(U) = \mathcal{Q}_P(BC). \tag{147}$$

Then every perturbation ΔC satisfying

$$0 < \|B\Delta C\|_\infty < \rho_{\mathcal{Q}_P}(BC) \tag{148}$$

produces a distinct fitted prediction with the same field d. When $r > 0$ and the admissible coefficient set contains a neighborhood of C relative to the attainable coefficient subspace, such a perturbation can be made arbitrarily small by scaling a nonzero attainable direction. This argument is conditional on Equation (147) and does not guarantee a second candidate for an arbitrary transform or constrained domain.

8.4 Unknown basis and structural metadata

Possession of both numerical arrays C and R need not be sufficient when the basis is unknown. Let

$$B_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad B_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad C = [1], \qquad R = 0. \tag{149}$$

Both bases are orthonormal, but they yield different originals:

$$B_1 C + R = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad B_2 C + R = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{150}$$

Unknown segment boundaries, starts, row ordering, feature ordering, or inverse scaling can create additional ambiguity even when the numerical segment values are otherwise known.

The exact general criterion is simple. Let θ denote all reconstruction metadata, let D_θ map fitting-domain values back to the source domain, and let $O(\theta, Y)$ be the released observation. Exact reconstruction without θ is possible if and only if

$$O(\theta, Y) = O(\theta', Y') \quad \Longrightarrow \quad D_\theta(Y) = D_{\theta'}(Y') \tag{151}$$

for every admissible pair. Missing metadata is not automatically ambiguous, but Equation (151) states exactly what a proof must establish.

9 Affine Ambiguity from Withheld Location-Scale Metadata

Consider a preprocessing map applied independently to each segment and column. Suppressing those indices, write

$$\mathcal{N}(x) = \frac{x - \mu(x)\mathbf{1}}{\tau(x)}, \qquad \tau(x) > 0, \tag{152}$$

where μ is a location functional and τ is a positive scale functional. Assume that, on the stated operating domain, they satisfy positive affine equivariance:

$$\mu(ax + b\mathbf{1}) = a\mu(x) + b, \tag{153}$$

$$\tau(ax + b\mathbf{1}) = a\tau(x) \tag{154}$$

for every $a > 0$ and $b \in \mathbb{R}$.

Proposition 9.1 (Positive affine invariance)**.** *Fix segment boundaries and suppose Equation (154) holds for every transformed column under consideration. For positive per-column scales $a_j > 0$ and arbitrary offsets b_j, define*

$$X'_{ij} = a_j X_{ij} + b_j. \tag{155}$$

Then the normalized matrices for X and X' are identical in exact arithmetic.

Proof. For each column j, positive affine equivariance gives

$$\mathcal{N}(a_j X_{\cdot j} + b_j\mathbf{1}) = \frac{a_j X_{\cdot j} + b_j\mathbf{1} - (a_j\mu(X_{\cdot j}) + b_j)\mathbf{1}}{a_j\tau(X_{\cdot j})} \tag{156}$$

$$= \frac{X_{\cdot j} - \mu(X_{\cdot j})\mathbf{1}}{\tau(X_{\cdot j})} = \mathcal{N}(X_{\cdot j}). \tag{157}$$

The result follows column by column. □

Under Proposition 9.1, every deterministic downstream output derived solely from the normalized matrix is unchanged, including the fitted coefficients and residuals. The inverse location and scale metadata encode the discarded per-column offsets and positive scales.

Suppose the admissible source domain contains a distinct, nonidentity positive per-column affine transform X' of X, and every other released field is invariant under that transform. If the inverse location and scale metadata are withheld, then X and X' produce the same complete observation and the source array is not identifiable from that observation. The result also covers deterministic downstream selection or state construction because the complete normalized inputs are identical. It assumes fixed segment boundaries and transformations that remain in the admissible numerical domain. No claim is made for degenerate or exceptional preprocessing paths that fail Equation (154).

10 Constructive Evidence

10.1 An exact, well-conditioned witness

Use

$$B = \begin{bmatrix} -1 & 1 \\ -\frac{1}{2} & 1 \\ 0 & 1 \\ \frac{1}{2} & 1 \\ 1 & 1 \end{bmatrix}, \qquad z = \begin{bmatrix} 1 \\ -2 \\ 2 \\ -2 \\ 1 \end{bmatrix}. \tag{158}$$

All entries are exactly representable in common binary floating-point formats, and direct calculation gives

$$B^T B = \begin{bmatrix} \frac{5}{2} & 0 \\ 0 & 5 \end{bmatrix}, \qquad B^T z = 0, \qquad \kappa_2(B) = \sqrt{2}. \tag{159}$$

The example is deliberately well conditioned so that numerical instability cannot explain the collision.

10.2 Coefficient-only collision

Let

$$C = \begin{bmatrix} 10 \\ 20 \end{bmatrix}, \qquad BC = \begin{bmatrix} 10 \\ 15 \\ 20 \\ 25 \\ 30 \end{bmatrix}. \tag{160}$$

Define

$$Y_+ = BC + z = \begin{bmatrix} 11 \\ 13 \\ 22 \\ 23 \\ 31 \end{bmatrix}, \qquad Y_- = BC - z = \begin{bmatrix} 9 \\ 17 \\ 18 \\ 27 \\ 29 \end{bmatrix}. \tag{161}$$

Both are bounded, strictly increasing integer sequences. They have exactly the same least-squares coefficients C and the same fitted prediction BC. Their residuals have the same exact magnitude,

$$\|Y_+ - BC\|_F = \|Y_- - BC\|_F = \sqrt{14}. \tag{162}$$

They therefore agree under every deterministic norm-only auxiliary field. Because their fitted prediction is identical, they also agree under every deterministic field derived solely from that prediction. Their exact originals differ in every row.

More generally, every

$$Y_\lambda = BC + \lambda z \tag{163}$$

has the same coefficients for real λ. The exact-real fiber is infinite. Any finite-precision state space is finite, but the witness contains demonstrably distinct exactly representable members.

10.3 Residual-only collision

Fix $R = z$ and choose

$$C_1 = \begin{bmatrix} 10 \\ 20 \end{bmatrix}, \qquad C_2 = \begin{bmatrix} 12 \\ 18 \end{bmatrix}. \tag{164}$$

Then

$$Y_1 = BC_1 + R = \begin{bmatrix} 11 \\ 13 \\ 22 \\ 23 \\ 31 \end{bmatrix}, \qquad Y_2 = BC_2 + R = \begin{bmatrix} 7 \\ 10 \\ 20 \\ 22 \\ 31 \end{bmatrix}. \tag{165}$$

Both are bounded, strictly increasing integer sequences and produce exactly the same residual R, and therefore the same residual-magnitude field. Their fitted predictions differ. Whether an additional deterministic prediction field also agrees depends on its transform; this witness alone does not establish a collision for the joint observation (R, d).

Proposition 10.1 (Exact bounded residual-fiber size)**.** *For the witness B and z in Equation (158), define*

$$\mathcal{S}_{\uparrow,31} = \left\{ y \in \mathbb{Z}^5 : 0 \le y_1 < y_2 < y_3 < y_4 < y_5 \le 31 \right\}. \tag{166}$$

Then

$$|\{ y \in \mathcal{S}_{\uparrow,31} : Qy = z \}| = 24. \tag{167}$$

Proof. Appendix A gives an exhaustive integer parametrization and the closed-form count $12 + 8 + 4 = 24$. □

10.4 Exhaustive bounded-domain enumeration

The witnesses above are not isolated continuous-space artifacts. Exact enumeration produced the results in Table 3. For this linear basis, coefficient equality on integer vectors is determined by equality of the sum and the weighted sum with weights $(-2,-1,0,1,2)$. Residual equality is determined by equality of the three consecutive second differences

$$s(y) = (y_1 - 2y_2 + y_3, y_2 - 2y_3 + y_4, y_3 - 2y_4 + y_5)\,, \tag{168}$$

because $s(y) = s(y')$ exactly when $y - y'$ is an affine sequence, equivalently when $y - y' \in \mathrm{col}(B)$.

Table 3: Exact enumeration of bounded integer component fibers. Each row states the observed component.

Admissible domain	Inputs N	Observations m	Singleton fibers	Inputs in non-singleton fibers	Largest fiber
Coeff.: $\{0,\dots,7\}^5$	32,768	1,016	104	32,664	118
Coeff.: nondecreasing in $\{0,\dots,31\}^5$	376,992	9,210	858	376,134	172
Residual: strictly increasing in $\{0,\dots,31\}^5$	201,376	13,915	1,460	199,916	112

Applying Corollary 5.4 and the complete exact fiber histograms gives Table 4. Every bit value is conditional uncertainty or information under the specified uniform prior.

Table 4: Uniform-prior exact-recovery and information measures for the bounded component observations. Candidate-set statistics are weighted by a uniformly random admissible input. The exact recovery ratios are 1,016/32,768, 9,210/376,992, and 13,915/201,376.

Admissible domain	P_{guess}	$\widetilde{H}_\infty(Y \mid V)$	$H(Y \mid V)$	$I(Y;V)$	$\mathbb{E}[K]$
Coeff.: $\{0,\dots,7\}^5$	3.100586%	5.011315	5.729947	9.270053	65.043945
Coeff.: nondecreasing in $\{0,\dots,31\}^5$	2.443023%	5.355189	6.175354	12.348821	90.939553
Residual: strictly increasing in $\{0,\dots,31\}^5$	6.909959%	3.855179	4.596393	13.023140	32.702686

For comparison, Table 5 evaluates the same priors when the complete cumulative view is constant. This is the exact prior-only baseline from Corollary 5.8; it is not obtained merely by deleting the coefficient and residual arrays from a package that retains data-dependent metadata.

Table 5: Uniform-prior baseline when the complete cumulative observation is constant. The exact recovery ratios are 1/32,768, 1/376,992, and 1/201,376.

Admissible domain	P_{guess}	$\widetilde{H}_\infty(Y \mid V_\varnothing)$	$H(Y \mid V_\varnothing)$	$I(Y;V_\varnothing)$	$\mathbb{E}[K]$
$\{0,\dots,7\}^5$	0.003051758%	15.000000	15.000000	0	32,768
Nondecreasing in $\{0,\dots,31\}^5$	0.000265258%	18.524174	18.524174	0	376,992
Strictly increasing in $\{0,\dots,31\}^5$	0.000496584%	17.619532	17.619532	0	201,376

For each domain, observing a deterministic component multiplies the constant-view exact-recovery probability by the number m of attainable component observations in Table 3. Table 6 shows why no-array and no-information are different claims. It evaluates the separately declared decoded metadata view V_M^* from Equation (141).

Table 6: Uniform-prior measures for the configuration-scoped decoded metadata equation replay. The exact recovery ratios are 14,355/32,768, 243,260/376,992, and 139,336/201,376. The recorded MLX 0.31.2 runtime is outside the source tag's declared range; this is not release-conformance evidence.

Admissible domain	P_{guess}	$\widetilde{H}_\infty(Y \mid V_M^*)$	$H(Y \mid V_M^*)$	$I(Y;V_M^*)$	$\mathbb{E}[K]$
$\{0,\dots,7\}^5$	43.807983%	1.190734	1.586127	13.413873	3.971130
Nondecreasing in $\{0,\dots,31\}^5$	64.526568%	0.632035	0.868468	17.655707	2.264128
Strictly increasing in $\{0,\dots,31\}^5$	69.191959%	0.531324	0.732913	16.886620	2.006654

The metadata values are exact finite-domain census results for the frozen equality rule, with only the displayed logarithms rounded. The three metadata maps have respectively 6,962, 162,546, and 99,162 singleton fibers,

and largest fibers of size 20, 20, and 15. The digest alone induces 1,016, 9,210, and 6,942 attainable values; the magnitude alone induces 649, 5,129, and 4,094. Their joint intersections induce the much larger counts in the table caption. These censuses do not estimate a natural-population attack rate. Generic hash preimage hardness does not prevent exhaustive lookup over a small declared source domain.

The full histograms also determine the input-weighted candidate-set distribution summarized in Table 7. Percentages are rounded only for display.

Table 7: Probability that a uniformly random admissible input lies in a component fiber whose size K falls in the stated band.

Admissible domain	$K = 1$	2–4	5–9	10–19	20–49	50–99	$K \geq 100$
Coeff.: $\{0, \ldots, 7\}^5$	0.317%	1.044%	2.594%	6.311%	23.193%	48.645%	17.896%
Coeff.: nondecreasing in $\{0, \ldots, 31\}^5$	0.228%	0.923%	1.942%	4.491%	16.627%	30.351%	45.439%
Residual: strictly increasing in $\{0, \ldots, 31\}^5$	0.725%	5.502%	7.012%	20.372%	42.011%	24.114%	0.264%

Using $q_\alpha = \min\{k : \Pr[K \leq k] \geq \alpha\}$, the median candidate-set sizes for a random admissible input are 67 and 93 for the coefficient domains and 28 for the residual domain. For a uniformly selected attainable observation, the corresponding medians are only 20, 22, and 8. This contrast is the size bias proved in Proposition 5.7.

Table 8 strengthens the selected-fiber counts with exact approximate-reconstruction and squared-error quantities. Each conditional posterior is uniform because the corresponding domain-wide prior is uniform and the observation map is deterministic.

Table 8: Observation-conditioned reconstruction quantities for three fully enumerated fibers. Distances are Euclidean, equivalently Frobenius for these one-column vectors, and squared-error risks are total rather than per-entry. The strict-radius column is the exact optimum for one returned reconstruction under the stated uniform conditional posterior. The Bayes-risk decoder may return any point in the ambient Euclidean fitting space; restricting its output to admissible originals can only increase that risk. The final column is the Corollary 7.3 lower bound on fiber-conditional worst-case expected squared error, $\mathrm{diam}^2/4$, and does not lower-bound the adjacent Bayes risk. Additional metadata, releases, or side information require a new joint fiber and posterior.

Selected component fiber	K	$d_{\min}$	$P^*\{d < d_{\min}/2\}$	R^*_{SE}	Fiber minimax SE lower bound
Coefficient, cube	55	2	1/55	$8728/605 \approx 14.4264$	63/2
Coefficient, nondecreasing	105	2	1/105	$107558/3675 \approx 29.2675$	114
Residual, strictly increasing	24	$\sqrt{5}$	1/24	$1625/36 \approx 45.1389$	605/4

The corresponding pointwise conditional min-entropies are $\log_2 55 = 5.781360$, $\log_2 105 = 6.714246$, and $\log_2 24 = 4.584963$ bits. These values concern the selected observations, not the average observation in each domain.

These counts show that the stated structural constraints do not automatically restore uniqueness on the three enumerated domains. Their broader scope is limited as described in Section 14.

11 Generalized Implementation Conformance

The controlled conformance framework evaluates whether a fixed implementation realizes the reference model through finite-precision segment fitting, optional normalization that is invertible given retained per-feature metadata, sufficient reconstruction metadata, auxiliary fields, and any serialization transforms. A logical decoded metadata view may be written $V_M = (H, M)$, where H contains in-scope top-level structure and M contains every segment-level field remaining after both numerical arrays are removed. A serialized or operational view may contain additional fields and side channels, so equality of V_M does not establish equality of the complete package view. This edition does not itself establish conformance for a particular build. Exact preprocessing parameters, numerical routines, metadata schemas, basis-construction details, and serialization encodings remain in the controlled record.

More generally, an implementation falls within the analyzed class when its observed behavior can be represented as follows:

1. data are divided into segments whose placement is recoverable from structural metadata;
2. an optional segment-local, per-feature location-scale transform is applied and later inverted using retained metadata;
3. each fitting-domain segment uses a fixed or reconstructible basis B;
4. the fitted coefficients, prediction, and residual satisfy $C \approx B^{+}Y$, $P \approx BC$, and $R \approx Y - P$ under declared finite-precision tolerances;
5. full fitting-domain reconstruction recombines P and R, after which any source-domain inverse transform is applied; and
6. every released auxiliary field is included in the observer's complete observation map; and
7. the observation map includes the ordered history of all releases and disclosed certificates available to the observer, rather than only the current artifact.

The exact theorem describes the ideal real-arithmetic reference model. Conformance tests must separately establish tolerance-level or bit-level collisions for the in-scope finite-precision realization. Table 6 is a source-equation replay of one frozen decoded metadata map. It is evidence for that map only, not evidence that a complete token, distributed build, or cumulative release history factors through it.

11.1 Reconstruction metadata categories

Table 9 identifies the metadata categories material to reconstruction without exposing schema names or encodings.

Table 9: Generalized reconstruction metadata and security-relevant observations.

Metadata category	Role in reconstruction	Security or correctness observation
Shape and segment placement	Places each reconstructed segment	Reveals dimensions, counts, boundaries, and potentially change points
Basis description	Reconstructs an analytical or configured basis	Must include every parameter on which B depends
Data-derived reconstruction state	Reconstructs a basis or other state fitted from data	May itself reveal information and can be larger than ordinary metadata
Inverse location and scale	Returns fitting-domain values to source units	Withholding it preserves the affine ambiguity only on the domain of Proposition 9.1
Source-unit conversion metadata	Interprets intermediate values in source units	Cannot recover information already discarded by an earlier conversion
Coefficient payload	Supplies the fitted component	Reveals basis-captured structure such as level and trend
Residual payload	Supplies the complementary component	May reveal events, local differences, noise signatures, or high-frequency structure
Residual-encoding metadata	Interprets a serialized residual payload	Quantization can change exact-equality and losslessness claims
Residual-magnitude field	Reports a scalar function of residual energy	Narrows coefficient fibers; it is redundant beside the exact pre-serialization residual when it is a known function of that residual
Conditional prediction field	Represents a deterministic function of the fitted prediction	May verify guesses and narrow residual-side fibers

Analytical, configured, and data-dependent bases can require different metadata. A split-component protocol must therefore analyze the complete basis description actually available to the observer rather than assuming that a short basis label is sufficient.

11.2 Serialized residuals

A serialized representation may apply a deterministic quantizer $\mathcal{Q}_R$ to the residual:

$$\widetilde{R} = \mathcal{Q}_R(R). \tag{169}$$

Consequently:

- Theorem 3.1 applies directly to the pre-serialization residual.
- A claim about serialized residual values requires a quantization-aware collision test.
- Exact component validation should use an unquantized or in-memory observation path.
- "Original" must distinguish source-domain data, fitting-domain values, and any intermediate representation.

12 Security Interpretation

The coefficient and residual components are deterministic, semantically meaningful projections. They are not cryptographic shares. Proposition 5.5 gives the exact statement: $I(Y;V) = H(V)$, so the mutual information is positive if and only if the induced view V is not almost surely constant under the declared prior. Low exact-recovery probability and nonzero information leakage can therefore occur at the same time, as Table 4 demonstrates. Mutual information does not by itself identify whether the revealed content is sensitive, but any nonconstant deterministic view under a full-support prior fails Shannon perfect secrecy about the complete input [4].

Coefficients may reveal smooth trends, offsets, periodic structure, or nearly the entire signal when the residual is small. Residuals may reveal sample-level differences, anomalies, high-frequency content, subject-specific noise, or event timing. With a mean-only basis, for example,

$$R_i = Y_i - \bar{Y} \tag{170}$$

and therefore

$$R_i - R_j = Y_i - Y_j. \tag{171}$$

The residual removes the mean coordinate algebraically but preserves every pairwise difference. It does not necessarily conceal the mean under a correlated prior. Knowledge of one source value or the mean restores the entire segment.

The defensible description of the present mechanism is *controlled selective disclosure*. A policy selects the complete view released to a recipient, including the component, basis information, precision, reconstruction metadata, protocol transcript, and prior release history. Theorem 6.2 makes the control precise for nested views: adding observable fields shrinks or preserves every fiber and cannot reduce an optimal observer's reconstruction power. Theorem 6.4 applies the same analysis when complementary channels are administered by different roles.

There is not, in general, one universal scalar disclosure level. Coefficient-only and residual-only views reveal different coordinates and semantic content, and neither view is necessarily a deterministic function of the other. They must therefore be compared under a declared domain, prior, target, metric, and complete metadata view. Basis choice, rank, precision, channel allocation, and metadata policy are disclosure design variables, but their quantitative effects must be computed rather than assumed.

Within the exact reference model, a receiver's view can be non-injective while a larger designated coalition is reconstruction-complete. This characterizes the declared channel assignment. It does not prove authorization enforcement, and it is not cryptographic secret sharing. Authentication, confidentiality in transit and at rest, resistance to unauthorized collusion, key management, policy enforcement, and secure deletion require separate mechanisms and evidence. A public smart contract may record commitments, enforce public conditions, verify attestations, or coordinate encrypted release; it cannot conceal plaintext state, calldata, events, transaction fields, or a key published on chain.

Proposition 12.1 (An unrestricted compatible candidate is easy to construct)**.** *Assume the basis B is known and the fitting domain is unrestricted. For every attainable coefficient observation C, the matrix BC is a compatible input. For every attainable residual observation R, the matrix R is a compatible input. Both candidates are obtained in polynomial time in the matrix dimensions.*

Proof. If $C = B^+Y$ is attainable, the Moore–Penrose identity $B^+BB^+ = B^+$ gives $B^+(BC) = B^+BB^+Y = C$. If $R = QY$ is attainable, idempotence of the projector gives $QR = Q^2Y = R$. The constructions require only ordinary matrix operations. □

Proposition 12.1 rules out using fiber multiplicity as evidence that finding *some* unrestricted compatible input is hard. Any computational claim must instead define a constrained, likely-source, attribute, or other search problem with an explicit input encoding and resource parameter, or rely on a separate cryptographic construction.

Definition 12.2 (Differential privacy for a declared adjacency)**.** Let $\sim$ be a declared adjacency relation on $\mathcal{S}$. A possibly randomized mechanism M is (ε, δ)-differentially private if, for every adjacent pair $y \sim y'$ and every measurable output event A,

$$\Pr[M(y) \in A] \le e^{\varepsilon} \Pr[M(y') \in A] + \delta. \tag{172}$$

The probabilities are over the mechanism's internal randomness [5].

Proposition 12.3 (Deterministic differential-privacy boundary)**.** *Assume distinct output values can be separated by a measurable event, as in every finite, discrete, or canonically serialized output space. Let $M : \mathcal{S} \to \mathcal{O}$ be deterministic, let $\varepsilon \ge 0$, and let $0 \le \delta < 1$. Then M is (ε, δ)-differentially private if and only if*

$$M(y) = M(y') \qquad \text{for every } y \sim y'. \tag{173}$$

Equivalently, M must be constant on each connected component of the adjacency graph. If the graph is connected, M must be constant on $\mathcal{S}$.

Proof. If some adjacent pair has different outputs, choose a measurable event A containing $M(y)$ but not $M(y')$. Determinism gives

$$1 = \Pr[M(y) \in A] \le e^{\varepsilon} \Pr[M(y') \in A] + \delta = \delta. \tag{174}$$

This is impossible when $\delta < 1$. Conversely, if outputs agree on every adjacent pair, their output distributions are identical, so the mechanism is $(0, 0)$-differentially private and therefore also satisfies every weaker (ε, δ) guarantee. Equality propagates along every adjacency path. □

Fiber multiplicity alone does not imply differential privacy. A future distorted release must declare the privacy unit and neighboring relation and prove the (ε, δ) inequality for the complete output distribution, including metadata and repeated-release composition [5]. Calibrated randomization is the standard nontrivial construction route; Proposition 12.3 states the precise obstruction for the present deterministic case.

13 Validation Protocol

A reviewer-grade evidence package should separate proof, numerical corroboration, serialization behavior, and privacy measurement.

13.1 Symbolic witness verification

Using exact rational arithmetic, verify:

$$B^TB = \operatorname{diag}(5/2, 5), \qquad B^Tz = 0, \tag{175}$$

the ranks of B^+ and Q, and the fiber dimensions from Theorem 3.1. These are exact checks for the displayed witnesses; the general theorem is established by its proof in Section 3. Numerical tests are corroboration.

Appendix B governs solver and serialization conformance, including numeric environments, tolerances, canonical equality, and collision criteria. It distinguishes numerical corroboration from an exact implementation collision.

13.2 Domain enumeration and adversarial inference

Preserve the bounded enumeration in Table 3, the constant-view baselines in Table 5, the metadata-only results in Table 6, the selected-fiber quantities in Table 8, and the complete fiber histograms as deterministic regression artifacts. Exact enumeration is a census of the declared finite domain, so sampling confidence intervals are neither needed nor appropriate for these tables or Table 7. Every change to the domain, prior, complete cumulative observation, side information, metric, success radius, list size, loss, quantization, runtime, or equality rule requires recomputation.

The companion v0.1 ambiguity envelopes certify only candidate-multiplicity lower bounds for three selected observations: at least 55, 105, and 24 admissible originals under their declared constraints. The exhaustive census establishes exact equality for those selected fibers, their minimum separations, squared-error quantities, and the domain-wide entropy calculations under the stated uniform priors. The envelope schema and reference verifier do not evaluate operational package factorization, release-history completeness, issuer authenticity, entropy, hardness, differential privacy, approximate-reconstruction radius, or inference resistance. An envelope that discloses explicit witnesses can also reduce the search needed to find compatible candidates even though a canonically generated envelope does not refine the declared information fiber.

Distribution-specific practical inference is a separate experiment. Evaluate explicit attacks under at least three threat models: component only; component plus all reconstruction metadata; and component plus metadata and population priors. Split by independent subject or entity when repeated records would otherwise leak across train and test sets. Freeze preprocessing and attack selection before evaluating the held-out set.

Recommended attacks include coarse reconstruction BC, learned residual or coefficient prediction, constraint solving, nearest-neighbor identification, sensitive-attribute classification, waveform or event classification, and membership inference. Useful metrics include exact recovery rate, candidate count, $\log_2$ candidate count, RMSE, R^2, correlation, spectral coherence, event accuracy, identification accuracy, and AUC. Report held-out estimates with confidence intervals and compare against prior-only, random, and full-input controls.

These model-, data-, and budget-dependent experiments address the broader claims bounded in Section 14; they neither prove nor refute the fiber theorems.

13.3 Acceptance criteria

The expanded claim is supported only if:

1. the exact symbolic identities and fiber dimensions pass;
2. the bounded integer and monotone witnesses pass;
3. the complete cumulative deterministic view V, including ordered prior releases, metadata, and disclosed certificates, is frozen, and any correlated non-deterministic side information is specified through its joint law with Y;
4. the finite domain $\mathcal{S}$, prior π, and observer information are declared beside every probability;
5. the exact histogram n_j satisfies $\sum_j n_j = m$, $\sum_j j n_j = N$, and $\sum_{j\geq 2} j n_j$ equals the reported non-singleton-input count; for a constant view, $n_N = 1$ and all other n_j vanish;
6. every repeated fixed-linear-release claim reproduces the stacked operator $A_{1:T}$, its rank, and the kernel intersection in Theorem 4.2; adaptive or nonlinear histories are evaluated through their complete joint observation map instead;
7. P_{guess} is reproduced from Equation (38) and, for each uniform case, independently as m/N;
8. the identities $2^{-\widetilde{H}_\infty} = P_{\text{guess}}$ and $H(Y) = H(Y \mid V) + I(Y;V)$ pass, together with $I(Y;V) = H(V)$ when the view is deterministic;
9. an independent exact-integer implementation reproduces the ideal component histograms and all reported quantities; the metadata replay separately reproduces every observation key with the scalar one-column operation shapes and preserves a digest of the complete canonical observation census;

10. selected 55-, 105-, and 24-candidate results are labeled as observation-conditioned values rather than domain-wide averages, and an independent exact-integer implementation reproduces their minimum squared distances, exact Bayes total squared-error fractions, squared diameters, and diameter-derived lower bounds;
11. every approximate or list-reconstruction statement declares the metric, strict or closed success event, radius, candidate support, conditional prior, and maximum list size;
12. every Gaussian-prior statement declares the fixed operator, mean, variance, rank, nullity, and whether loss is total squared error or per-entry mean squared error, then reproduces the posterior covariance trace and the stated chi-squared probability;
13. every party's stored, received, and publicly emitted fields are frozen, and the complete cumulative view of every in-scope coalition is enumerated;
14. each coalition designated to reconstruct satisfies the stated target-factorization or injectivity condition; a per-token ambiguity claim uses the actual observed fiber, a domain-wide non-reconstructability claim checks every attainable fiber, and a quantitative claim reports the declared Bayes or minimax bound rather than relying on one non-singleton example;
15. contract state, calldata, events, transaction metadata, commitments, ciphertexts, release decisions, and earlier protocol messages are assigned to the applicable public or coalition view; and
16. the evidence distinguishes access-policy enforcement from confidentiality, cryptographic sharing, secure deletion, and resistance to party compromise; and
17. external language conforms to the claim boundary, and every build-specific statement also satisfies Appendix B. Tolerance-level agreement is labeled as numerical corroboration, never as an exact implementation collision. A release-conformance claim additionally requires a supported dependency environment and bit-identical equality for the complete serialized observation.

14 Limitations and Counterexamples

Any external claim should state the following boundaries explicitly.

- **Identity basis.** If B is square and invertible, especially $B = I_L$, coefficients identify $Y = BC$. Any available identity basis is a direct counterexample to an unconditional coefficient claim.
- **Zero-rank basis.** If $r = 0$, then $R = Y$ and residuals identify the fitting-domain data.
- **Zero residual.** If a released magnitude certifies $\|R\|_F = 0$ and B, C, and inverse preprocessing are known, coefficients reconstruct exactly.
- **Constrained singleton fibers.** Bounds or physical constraints may reduce a particular fiber to one admissible element.
- **Data-dependent bases.** The fixed-B theorem does not by itself prove an end-to-end collision when changing Y also changes a data-dependent basis or an automatic selection outcome. Proposition 9.1, on its stated domain, avoids this issue by preserving the complete normalized input.
- **Degenerate or non-equivariant preprocessing.** Proposition 9.1 applies only where the scale functional is positive and Equation (154) holds. It makes no claim for exceptional preprocessing paths that violate those conditions.
- **Data-dependent segmentation.** Segment boundaries may reveal change points and may change under source transformations. The affine result assumes fixed boundaries.
- **Finite precision.** Exact-real fibers may be infinite, while every finite-precision realization has a finite state space. Evidence for such a realization must demonstrate actual distinct candidates under its chosen arithmetic and supported runtime. The metadata census used MLX 0.31.2, outside the referenced tag's declared MLX range, and is therefore an equation replay rather than release-conformance evidence.

- **Prior and support dependence.** The probabilities in Tables 4–6 assume a uniform prior on every member of the stated finite domain. Natural populations need not be uniform. Zero-probability candidates must be removed from the support, and a different prior can materially change every result.
- **Selected-view control is not a universal leakage budget.** A policy can control which declared view it releases. The amount and kind of information disclosed depend on the basis, precision, metadata, prior, side information, and target. Theorem 6.2 orders nested views, but non-nested views require separate analysis.
- **Complete-view, coalition, and composition dependence.** Additional deterministic fields refine the fiber partition. Correlated non-deterministic side information instead updates posterior probabilities through the joint law in Remark 5.3. The complete view includes ordered earlier releases, disclosed certificates, and every channel available to the declared coalition. Theorems 4.2 and 6.4 compute intersections for their stated maps. A guarantee for the receiver alone does not survive acquisition of a custodian's artifact, an originator's inverse metadata, another party's side information, or a later release unless the resulting coalition view is analyzed.
- **Originator knowledge and control-plane scope.** The data originator ordinarily knew the source before decomposition. Split custody controls downstream dissemination and recombination; it is not evidence that prior copies were deleted or that the originator cannot reconstruct. The fiber theorems also do not prove availability, revocation, fair exchange, correct policy execution, key custody, or resistance to a compromised administrator.
- **Smart-contract transparency.** Public contract state, calldata, events, transaction metadata, return values, and released keys are observable. A contract supplies confidentiality only when combined with a separately specified and verified encryption, threshold-key, confidential-execution, or trusted off-chain mechanism.
- **Metadata-replay scope.** Table 6 concerns only the declared decoded pair (D_P, S_R) under one fixed configuration. It excludes the complete binary token, file-wide residual-quantization scale, output length, timing, errors, adaptive choices, and other operational observables. Its values are not package-level leakage measurements.
- **No-extra-leakage factorization.** A claim that an operational package reveals no more than a declared leakage function requires a proved simulator or factorization for the complete package view and cumulative history. This paper does not establish that product guarantee.
- **Attestation disclosure.** The v0.1 ambiguity envelopes certify lower bounds for selected fibers, not domain-wide entropy. Canonical envelope generation need not refine the declared fiber, but publishing explicit witness candidates may make compatible originals easier to find. Issuer authentication and history completeness require separate mechanisms.
- **Exact versus approximate recovery.** Low exact-recovery probability alone does not imply large reconstruction error under squared, absolute, perceptual, or task-specific loss. Theorem 7.2 applies only after the required packing has been established for the stated metric and radius. Theorem 5.10 applies only under its declared Gaussian prior. Otherwise Equation (74) must be evaluated for the intended loss and population.
- **Attributes and identity.** Large fibers can have homogeneous sensitive attributes or identity signatures. Equation (101) and held-out attack results are required before making an attribute- or identification-resistance claim.
- **Continuous priors.** Under a non-atomic continuous prior, exact point recovery can have probability zero even when arbitrarily accurate approximation is easy. The finite quantization is part of the experiment, not a technical footnote.
- **Computational hardness.** Proposition 12.1 constructs an unrestricted fitting-domain preimage directly. Fiber multiplicity is therefore information-theoretic ambiguity, not evidence of computational one-wayness. A hardness claim requires a defined constrained, approximate, or high-prior-probability inverse problem and a reduction or serious complexity evidence.
- **Differential privacy.** Deterministic non-identifiability does not imply differential privacy. A claim requires a neighboring-data relation and proof of the (ε, δ) inequalities for the complete released view. Sensitivity-calibrated randomness is a standard construction, and composition must be analyzed when multiple releases occur.

15 Conclusion

The publication-ready statement is:

> **For a fixed basis B of rank r, under the standard least-squares decomposition, coefficients alone admit no universal exact inverse on the unrestricted real-valued input space when $r < L$, and residuals alone admit no universal exact inverse when $r > 0$. On a constrained input domain, exact reversibility is governed by whether that domain contains two distinct elements differing along the corresponding invisible subspace. For repeated fixed linear releases, the cumulative invisible subspace is the intersection of the individual kernels and has dimension $D(L - \operatorname{rank}(A_{1:T}))$. On the unrestricted fitting domain, cumulative injectivity occurs exactly when that intersection is zero; on a restricted domain, the corresponding difference-set condition must hold. For per-segment, per-column location-scale preprocessing that satisfies positive affine equivariance on its stated domain, if the source domain contains two distinct inputs related by such a transformation and every other released field is invariant, withholding inverse location and scale metadata makes those inputs observationally indistinguishable.**
>
> **A declared policy can choose which component and metadata view becomes observable and can distribute complementary reconstruction material across an originator, custodian, and receiver. For any coalition, the compatible-original set is the intersection of the fibers induced by that coalition's complete cumulative deterministic view. Nested disclosures can only shrink or preserve ambiguity and cannot worsen optimal reconstruction performance. In the ideal exact component model, coefficients alone leave the coefficient fiber; joining the exact pre-serialization residual and an available basis recovers the fitting-domain input as $Y = BC + R$. Fields retained by the originator may then support the intended source-domain inverse, but whether their absence prevents source recovery is a separate fiber question, and the originator is presumed already to know the source. This establishes policy-selected disclosure and reconstruction criteria for designated coalitions under the stated observation maps.**
>
> **On any finite admissible domain $\mathcal{S}$, under a declared prior π and complete deterministic observation $V = O(Y)$, the Bayes-optimal exact-recovery probability is $\sum_o \max_{y \in \mathcal{F}(o)} \pi(y)$. Under a uniform prior it equals $|O(\mathcal{S})|/|\mathcal{S}|$. If the complete cumulative view is constant, this reduces to $1/|\mathcal{S}|$ and $I(Y;V) = 0$; on the three declared domains the values are approximately 0.003052%, 0.000265%, and 0.000497%. The two enumerated coefficient views give 3.100586% and 2.443023%, and the enumerated residual view gives 6.909959%. In the selected 55-, 105-, and 24-member fibers, minimum separations 2, 2, and $\sqrt{5}$ make the optimal strict half-separation success probabilities $1/55$, $1/105$, and $1/24$ under their uniform conditional posteriors. Under an isotropic Gaussian prior, coefficient-only and residual-only Bayes total squared-error risks are exactly $\sigma^2(L-r)D$ and $\sigma^2 rD$, with optimal radius-success probabilities given by the corresponding chi-squared distribution functions. A separately scoped decoded metadata replay retaining a prediction digest and residual-magnitude field gives 43.807983%, 64.526568%, and 69.191959%.**
>
> **These results establish only the stated view-, metric-, domain-, prior-, and coalition-conditional reconstruction limits. They do not establish practical approximate-reconstruction resistance on representative natural populations; identification, linkage, membership, or attribute-inference resistance; computational one-wayness; encryption or perfect secrecy; differential privacy; release-history completeness; package-level no-extra-leakage factorization; cryptographic secret sharing; resistance to a custodian's disclosure or party collusion; or confidential public-smart-contract custody.**

A Reproducible Enumeration Procedure

For integer vectors $y = (y_1, \dots, y_5)$ under the witness basis, define the exact key

$$k(y) = \left(\sum_{i=1}^{5} y_i,\ -2y_1 - y_2 + y_4 + 2y_5 \right). \tag{176}$$

Two vectors have the same least-squares coefficients if and only if their keys agree. For every key value u, retain its exact fiber size

$$k_u = |\{y \in \mathcal{S} : k(y) = u\}|, \tag{177}$$

and then form the complete histogram

$$n_j = |\{u : k_u = j\}|. \tag{178}$$

The enumeration must verify

$$\sum_j n_j = m, \qquad \sum_j j n_j = N, \qquad \sum_{j \geq 2} j n_j = N_{\text{nonsingleton}}. \tag{179}$$

The bounded counts in Table 3 follow directly. No floating-point comparison is involved.

Compute the uniform-prior measures from Equations (44), (49), (52), (57), (64), and (65). The displayed decimals are rounded only after exact counting. Shannon entropy and expected candidate count require the complete histogram.

For the constant-view baseline, the observation map has one value on the entire domain. Its histogram is therefore $n_N = 1$ and $n_j = 0$ for $j \neq N$. This gives every entry in Table 5 analytically and supplies the simplest histogram-consistency check.

For the decoded metadata replay, use the degree-one Vandermonde basis with rows $(t_i, 1)$, where

$$(t_1, t_2, t_3, t_4, t_5) = (0, 1/4, 1/2, 3/4, 1). \tag{180}$$

Inputs are float32 with one feature, one segment, and preprocessing disabled. The frozen logical view retains shape $(5, 1)$, one fixed feature label, a null index, and one segment record. Its start, length, coefficient width, degree, source scale, preprocessing bounds and flag, basis identifier, and stored-basis field are constant; only the digest and residual magnitude vary with the source. The caller-specified basis and segmentation configuration are included as fixed out-of-band interpretation. For each source vector, independently solve the two triangular systems induced by the float32 Cholesky factor of $B^T B$, form $P = BC$ and $R = Y - P$, and define

$$D_P = \mathrm{SHA256}\big(\mathrm{leInt64}\big(\mathrm{roundEven}(10^6 P)\big)\big), \tag{181}$$

$$S_R = \mathrm{bits32}\left(\mathrm{float32}\left(\frac{\mathrm{norm32}(R)}{\sqrt{5}}\right)\right). \tag{182}$$

Here byte equality includes little-endian signed-integer prediction bytes and the exact unsigned 32-bit pattern of the decoded magnitude. Each input is logically a separate $(5, 1)$ solve. The evidence program checks the complete ordered batched key list against scalar one-column replay for every input, groups exact pairs (D_P, S_R), verifies the histogram identities, and hashes a canonical census of every observation and its multiplicity. The recorded environment used MLX 0.31.2 and therefore does not satisfy the referenced tag's declared MLX requirement $\geq 0.30.6, < 0.31$; repetition under a supported runtime is required before making a release-conformance claim.

For the complete residual-domain census, use the exact key $s(y)$ in Equation (168). The null space of the second-difference map is precisely the two-dimensional space of affine sequences, which equals $\mathrm{col}(B)$. The key therefore labels the exact ideal-residual fibers without floating-point comparison. Enumerate the $\binom{32}{5} = 201{,}376$ strictly increasing integer vectors in $\{0, \dots, 31\}^5$, form the complete key histogram, and verify the same three histogram identities above.

The 24-member fiber containing the residual z also has a closed-form count. Equation (159) gives $Qz = z$, and $Qy = z$ holds exactly when $y - z \in \ker(Q) = \mathrm{col}(B)$. Because B has full column rank, each compatible y has a unique representation $y = z + BC$. Write $C = (c_1, c_2)^T = (2a, b)^T$. Because y and z are integral, $BC = y - z$

must be integral. The center row of B forces $b \in \mathbb{Z}$, and either half-step row forces $a \in \mathbb{Z}$. Conversely, integer a and b make every entry of BC integral. Direct substitution gives

$$y = BC + z = (b - 2a + 1, b - a - 2, b + 2, b + a - 2, b + 2a + 1)^T . \tag{183}$$

Strict increase forces $a \geq 5$. Because the sequence is then increasing, the bounds $0 \leq y_i \leq 31$ reduce to

$$2a - 1 \leq b \leq 30 - 2a, \tag{184}$$

which forces $a \in \{5, 6, 7\}$. These three values permit respectively 12, 8, and 4 integer values of b. The fiber therefore contains exactly

$$12 + 8 + 4 = 24 \tag{185}$$

admissible originals. This proves the selected count independently of the full census.

The companion exact-integer program `tools/polygen_nonidentifiability_enumeration.py` emits the complete coefficient and residual histograms, selected fibers, observation census, exact ratios, derived information measures, minimum squared distances, squared diameters, exact Bayes total squared-error fractions, and diameter-derived worst-case lower bounds to `output/conformance/polygen_nonidentifiability_exact_enumeration.json`. It contains regression assertions for every value reported in Tables 3–7 and Table 8; no floating-point comparison determines fiber membership.

The companion decoded-metadata program `tools/polygen_no_component_metadata_enumeration.py` emits the constant-view and metadata-only results, complete histograms, canonical observation-census digests, environment record, and replay checks to `output/conformance/polygen_no_component_metadata_enumeration.json`. Its constant-view entries are analytical; its metadata entries are runtime-specific numerical censuses under Equation (182).

The v0.1 lower-bound envelope specification, closed reference schema, and verifier are recorded at `docs/validation/leakage_contract_certified_ambiguity_v0.1.md`, `docs/validation/polygen_certified_ambiguity_v0.1.schema.json`, and `tools/polygen_ambiguity_certificate.py`. The byte-stable generated bundle is indexed by `output/conformance/ambiguity_certificates/manifest.json`; its separate deterministic report records 27 tamper checks and eight public normalization vectors at `output/conformance/ambiguity_certificates/self_test_report.json`. The manifest intentionally records that self-test generation was not run within the manifest-producing step. The separate report has status `reference_self_test_pass`. Neither artifact evaluates the operational package-factorization profile.

B Controlled Conformance Checklist

Build-specific correspondence must be evaluated against a fixed revision in the controlled evidence package. The checklist below states the required verification obligations without exposing implementation details:

1. identify the in-scope revision, configuration, solver, supported numeric environment, serialization mode, complete metadata view, and ordered release history in the controlled record;
2. capture Y, B, C, P, and R, and verify $P \approx BC$, $Y \approx P + R$, and $B^T R \approx 0$ under predeclared absolute and relative tolerances, recording $\operatorname{rank}(B)$ and $\kappa_2(B)$;
3. label tolerance-level agreement only as numerical corroboration of the reference model. An implementation-level collision requires distinct inputs to produce bit-identical canonical observation keys or bit-identical serialized released observations under the equality rule available to the observer;
4. demonstrate such exact collisions separately for coefficient and residual observations, and distinguish each projected-component collision from a whole-token collision;
5. verify every auxiliary field from its actual inputs and analyze the complete joint observation released by the protocol; a decoded metadata collision does not substitute for equality of a richer serialized package view;
6. verify inversion of location-scale preprocessing, exercise every exceptional preprocessing and segment-length path, and test affine invariance only on its documented nondegenerate domain;

7. test reconstruction with complete metadata and the effect of withholding each metadata category claimed to be required, including data-derived reconstruction state;
8. test pre-serialization components and every serialization path included in the claim, recording quantization error and whether each canonical collision survives encoding and decoding;
9. repeat the witnesses and metadata censuses in every numeric environment included in the claim and report maximum component error, $\max |B^T z|$, canonical observation-census digests, and whether each comparison is bit-identical or tolerance-equal;
10. when ambiguity envelopes are disclosed, distinguish certified lower bounds from exact census results, verify deterministic regeneration and tamper rejection, and assess whether witness disclosure changes the attacker's search cost; and
11. exercise the counterexamples in Sections 4 and 14, including rank extremes, zero residual, singleton fibers, data-dependent selection changes, and segmentation changes.

The domain, prior, frozen-view, histogram, and information-measure gates are governed by Section 13.3 and are not repeated here.

References

[1] A. Ben-Israel and T. N. E. Greville, *Generalized Inverses: Theory and Applications*, 2nd ed., Springer, 2003.

[2] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed., Johns Hopkins University Press, 2013.

[3] Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith, "Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data," *SIAM Journal on Computing*, vol. 38, no. 1, pp. 97–139, 2008.

[4] C. E. Shannon, "Communication Theory of Secrecy Systems," *Bell System Technical Journal*, vol. 28, no. 4, pp. 656–715, 1949.

[5] C. Dwork and A. Roth, *The Algorithmic Foundations of Differential Privacy*, Foundations and Trends in Theoretical Computer Science, vol. 9, nos. 3–4, pp. 211–407, 2014.

[6] T. W. Anderson, "The Integral of a Symmetric Unimodal Function over a Symmetric Convex Set and Some Probability Inequalities," *Proceedings of the American Mathematical Society*, vol. 6, pp. 170–176, 1955.