Sparse Recursion for Holographic Steganography

Distributed Secret Fields Under Adaptive Computational Depth

Flyxion
Independent Researcher

2026

Abstract

Conventional steganography assigns payload symbols to fixed carrier locations. This local accounting is fragile: cropping removes symbols outright, compression corrupts selected coefficients, and an adversary can search for the statistical residue of a spatially concentrated embedding policy. We formalize **Sparse Recursive Holographic Steganography** (SRHS), an architecture in which a payload is first transformed into a redundant distributed field and then embedded through a short, adaptively selected sequence of carrier revisions. Under this construction every sufficiently informative view of the carrier carries degraded evidence about the entire payload, and no single carrier location is uniquely responsible for a fixed payload fragment. A recurrent controller with parameters shared across steps estimates local embedding capacity, modifies only a sparse active set of carrier regions, decodes the resulting intermediate carrier, and reallocates its remaining distortion budget accordingly. This separates representational reach, governed by recursion depth, from parameter count, which stays fixed.

SRHS replaces the standard one-pass map from message bits to carrier coefficients with three coupled operations: holographic payload projection, sparse carrier routing, and recursive residual correction. We state the system as constrained channel coding, give explicit assumptions under which partial-view recovery degrades gracefully rather than failing at fixed boundaries, and show why recursive correction weakly dominates one-shot embedding at equal distortion whenever the carrier response is uncertain in advance. We specify a complete training objective, a decoding procedure, a three-observer threat model, and a falsifiable experimental program with preregistered success and failure conditions. The claim advanced here is deliberately narrow: distributed payload evidence combined with sparse adaptive revision improves the achievable robustness-detectability-distortion frontier relative to localized or uniformly dense embedding, at matched budget. It is not a claim that channel capacity can be circumvented. If the stated hypotheses survive the proposed ablations, steganographic design shifts from the placement of secret symbols toward the synthesis of recoverable secret fields.

1 Introduction

Steganographic systems are conventionally organized around a spatial metaphor. A payload is partitioned into symbols; symbols are assigned to pixels, blocks, tokens, frames, or transform coefficients; a decoder later reads those assignments back. Learned systems relax the assignment but frequently retain its underlying logic: an encoder makes a single forward pass, distributes a finite distortion budget across the carrier, and produces a stego object whose hidden content is recovered from the resulting perturbation.

This organization produces three linked weaknesses. Locality produces brittle failure, since destroying the region assigned to a payload block destroys that block regardless of how much unused evidence remains elsewhere in the carrier. A one-pass encoder commits to its full modification before it can observe the realized consequences of any of it, and carrier channels are heterogeneous enough that post-embedding transformations are only partially predictable in advance. Dense modification, finally, wastes distortion on locations that contribute little reliable information while enlarging the statistical surface available to a detector.

The architecture proposed here replaces symbol placement with field construction. The payload is projected into a set of correlated measurements, each individually insufficient but jointly informative about the payload as a whole. These measurements are embedded into a sparse subset of carrier regions. After each embedding step the system decodes its own intermediate output, estimates the residual payload error, and selectively revises the carrier. Computation is therefore recursive, while modification remains sparse at every step.

The term *holographic* is used operationally rather than by analogy to physical optics: it denotes distributed recoverability, in which partial observations preserve degraded evidence about the whole message rather than exact evidence about isolated fragments of it. The term *recursive* likewise denotes a concrete recurrence with shared parameters and an explicit stopping rule, not literal self-similarity or unbounded depth. Neither term commits the architecture to any particular carrier modality.

This paper makes four contributions. It introduces a payload representation designed for whole-message recovery from partial carrier views. It introduces a sparse recurrent embedding controller that spends distortion only where an additional revision is expected to reduce decoding error without increasing distinguishability. It states robustness and improvement results under explicit, falsifiable assumptions rather than as informal claims. It defines an evaluation protocol constructed specifically to prevent robustness gains from being purchased through unreported distortion, payload reduction, secret-key leakage, or steganalysis weak enough to be uninformative.

2 Problem Formulation

Let a cover object be $x \in \mathcal{X}$, drawn from a declared cover distribution $\mathcal{D}_{\mathcal{X}}$; let a payload be $m \in \{0,1\}^L$; and let a secret key be k, drawn from a keyspace large enough to resist exhaustive search. An encoder produces

$$y = E_\theta(x, m, k; B),$$

where y is the stego object and B is a hard distortion budget. A channel transformation $T \sim \mathcal{T}$, drawn from a declared transformation distribution, may crop, compress, rescale, transcode, paraphrase, drop packets, reorder frames, or otherwise alter the object before it reaches the receiver. The decoder returns

$$\hat{m} = D_\phi\big(T(y), k\big).$$

A useful steganographic system must jointly minimize decoding error, perceptual or semantic distortion, and detectability. These three objectives conflict in general, so the target is Pareto improvement of the achievable frontier rather than a single universal optimum.

For a detector A, define its advantage under equal priors on cover and stego objects as

$$\mathrm{Adv}(A) = \big|\Pr[A(y) = 1] - \Pr[A(x) = 1]\big|.$$

The operational objective of SRHS is reliable payload recovery after transformations drawn from $\mathcal{T}$, subject to the hard distortion constraint $\|y - x\| \leq B$ under a declared metric, together with low advantage for both the detectors used during training and detectors unseen during training. Security in this sense is not established by perceptual similarity alone; it must be measured against adaptive steganalysis and, where the cover distribution is well characterized, bounded with respect to that distribution. Every result below is stated relative to a declared $(\mathcal{D}_\mathcal{X}, \mathcal{T}, B)$ triple, and no claim in this paper is intended to hold outside such a declaration.

3 Holographic Payload Projection

A localized representation maps payload block m_i primarily to a single carrier region j. SRHS instead maps the entire payload to a redundant collection of keyed measurements. An outer error-correcting code first produces $c = C(m) \in \{0,1\}^n$. Because c is a $\{0,1\}$-valued vector, projecting it directly carries a codeword-dependent DC component that inflates the measurement magnitude and complicates distance-preservation arguments. The codeword is therefore centered before projection,

$$\bar{c} = 2c - \mathbf{1},$$

so that $\bar{c} \in \{-1,+1\}^n$, and a keyed linear projection generates

$$z = \frac{1}{\sqrt{n}} P_k \bar{c}, \qquad P_k \in \mathbb{R}^{q \times n},$$

followed by a bounded quantization $u = Q(z)$. The $1/\sqrt{n}$ normalization keeps the expected measurement energy independent of n, which is what the restricted-isometry-type argument of Assumption 1 requires. The rows of P_k are generated from a cryptographic pseudorandom seed and normalized so that evidence is dispersed across measurements rather than concentrated in a few of them. Carrier regions each receive overlapping subsets of the measurements in u.

Assumption 1 (Restricted preservation)**.** *There exists a family of subsets $S \subseteq \{1, \ldots, q\}$, occurring with non-negligible probability under the transformations in $\mathcal{T}$, such that the restricted projection $P_{k,S}$ preserves the pairwise distances among codewords of C up to a bounded factor in each direction. This is a restricted-isometry-type condition on P_k in the sense of compressed sensing [3], applied here to a finite codeword set rather than to arbitrary sparse vectors. Concretely, for constants $0 < \gamma(S) \leq \Gamma(S)$ depending on S, and for any two codewords $c_a \neq c_b \in C$ with centered forms $\bar{c}_a, \bar{c}_b$,*

$$\gamma(S)\, \|\bar{c}_a - \bar{c}_b\|^2 \;\leq\; \|P_{k,S}(\bar{c}_a - \bar{c}_b)\|^2 \;\leq\; \Gamma(S)\, \|\bar{c}_a - \bar{c}_b\|^2.$$

The two directions of this bound serve different purposes and are not required to behave symmetrically as S varies. When $P_{k,S}$ is formed by restricting P_k to the rows indexed by S, adding rows can only add non-negative squared-norm contributions, since $\|P_{k,S'}v\|^2 = \|P_{k,S}v\|^2 + \|P_{k,S'\setminus S}v\|^2$ for $S \subseteq S'$. This gives $\gamma(S) \leq \gamma(S')$ whenever $S \subseteq S'$, so the lower gain γ is monotone non-decreasing in the surviving set under simple row restriction. The upper constant $\Gamma(S)$ has no corresponding guarantee: it can increase as rows are added, since the added rows also contribute to the squared norm of the ambient vector's image without bound in general. The upper bound is retained only to control embedding energy and distortion in Section 6; the recovery argument of Corollary 1 depends solely on the monotone lower gain $\gamma(S)$.

Corollary 1 (Decoding margin under restricted preservation)**.** *Let $d_{\min}$ denote the minimum Hamming distance of the outer code C. Since $\bar{c} = 2c - \mathbf{1}$, codewords at Hamming distance d satisfy $\|\bar{c}_a - \bar{c}_b\|^2 = 4d$, so the minimum squared distance among centered codewords is $4d_{\min}$. Under Assumption 1, the minimum squared distance among the normalized measurement images $z_a = \frac{1}{\sqrt{n}} P_{k,S}\bar{c}_a$ is at least $\gamma(S) \cdot 4d_{\min}/n$, and nearest-codeword decoding over the surviving measurements S succeeds whenever the aggregate noise, inclusive of quantization error, satisfies*

$$\|\eta_S\| \;<\; \sqrt{\frac{\gamma(S)\, d_{\min}}{n}}.$$

Because γ is monotone non-decreasing in S under row restriction, this threshold is monotone non-decreasing as S grows.

This margin makes the mechanism behind Proposition 1 explicit rather than qualitative, and it isolates exactly which half of the restricted-preservation bound the recovery argument needs: only $\gamma(S)$, not the symmetric distortion factor used informally above. It also gives a concrete design target for Section 9: $\gamma(S)$ can be estimated empirically on surviving subsets under the declared transformation distribution $\mathcal{T}$ and compared against $d_{\min}/n$, which is a falsifiable numerical claim rather than an appeal to the assumption's plausibility.

Under Assumption 1, the condition for recovery is not that every crop reconstruct the message exactly. It is that a sufficiently large family of measurement subsets retains enough restricted gain to support decoding. If S indexes the measurements surviving a given transformation, the decoder observes

$$u_S = \frac{1}{\sqrt{n}}\, P_{k,S}\, \bar{c} + \eta_S,$$

where η_S aggregates embedding noise, channel damage, internal decoding error, and quantization error. By Corollary 1, the sufficient recovery condition strengthens monotonically as S grows; Section 8.1 gives the corresponding statement for actual decoding risk, which requires a separate argument.

For multimedia carriers, a *view* is any transformation that preserves a subset or mixture of carrier evidence: a crop, a temporal window, a frequency band, a resolution level, or a set of surviving packets. Training samples multiple views of the same stego object and requires each to predict the same payload; this multi-view consistency is the learned counterpart of the projection's algebraic redundancy, and it is enforced explicitly in the training objective of Section 6.

The key controls the projection, the interleaving of measurements across regions, and the routing priors used in Section 4. It must not merely index a small public family of embeddings, since a small family is enumerable by an adversary. Key reuse and chosen-message attacks are accordingly treated as explicit components of the threat model in Section 10 rather than as out-of-scope concerns.

4 Sparse Recursive Embedding

Divide the carrier representation into N addressable regions, which may be spatial patches, wavelet bands, video tubes, audio time-frequency cells, text spans, or learned latent groups. At recursion step t the system maintains a stego candidate y_t, a residual payload state r_t, and a remaining budget b_t, initialized as

$$y_0 = x, \qquad r_0 = u, \qquad b_0 = B.$$

A controller with shared parameters ψ computes region scores

$$s_t = G_\psi\big(F(y_t),\, r_t,\, b_t,\, k\big),$$

where F is a carrier feature extractor. A sparse router selects an active set S_t of at most $K \ll N$ regions, using top-K selection, sparsemax, entmax, or a differentiable hard-concrete gate during training. A proposal network produces a bounded revision

$$\Delta_t = U_\theta\big(F(y_t),\, r_t,\, k,\, S_t\big),$$

and the carrier is updated only on the active set:

$$y_{t+1} = \Pi_{\mathcal{C}(x,B)}\big(y_t + M(S_t) \odot \Delta_t\big).$$

Here $M(S_t)$ is the active-region mask and $\Pi_{\mathcal{C}(x,B)}$ projects the candidate back into the admissible distortion set around x. Because the receiver's eventual observation is $T(y)$ for a transformation T sampled only after encoding is complete, the encoder cannot observe the realized future transformation and must instead probe with a surrogate drawn from the same declared distribution. The system samples $\widetilde{T}_t \sim \mathcal{T}$ and performs an internal decode against the probed carrier,

$$\tilde{u}_{t+1} = D_\phi^{\text{inner}}\big(\widetilde{T}_t(y_{t+1}), k\big), \qquad r_{t+1} = u - \tilde{u}_{t+1}.$$

This surrogate decode reveals information about carrier susceptibility under the transformation family $\mathcal{T}$, and about the rendering, quantization, and projection effects realized in y_{t+1} itself; it does not reveal which specific transformation the eventual receiver's channel will apply. What it can inform the next step's allocation about is therefore carrier response that is persistent across steps, such as which regions are structurally fragile under the kinds of damage in $\mathcal{T}$, rather than the identity of a one-time future draw. Section 8.2 makes this distinction precise, since it determines whether recursion has anything to learn from step to step.

Recursion terminates when the estimated robust decoding loss falls below a threshold, when the budget is exhausted, or when a fixed maximum depth R is reached. Because G_ψ, U_θ, and the internal decoder share parameters across steps, increasing depth increases adaptive computation without increasing parameter count.

This recurrence is not incidental to the architecture. The true response of a carrier region, once quantized, compressed, or otherwise passed through a differentiable channel approximation, is not knowable in advance of a rendered modification. Each step is simultaneously an intervention and a measurement: the encoder does not merely add signal to the carrier, it learns from a probed surrogate of what a receiver's channel could do to its preceding attempt.

5 Sparsity and Recursion as a Joint Requirement

Sparsity without recursion commits irrevocably to a single, possibly mistaken, active set. Recursion without sparsity perturbs the entire carrier at every step and compounds detectable residue across steps. Their conjunction permits exploration followed by correction: early steps allocate weak measurements to regions predicted to be safe, an intermediate decode reveals which payload directions remain uncertain, and later steps address those directions through new regions or revisions of unreliable ones.

Definition 1 (Marginal routing value)**.** *Let ℓ_t denote expected payload loss after step t and let $c_{t,j}$ denote the cost of revising region j at step t. The marginal value of revising region j is*

$$v_{t,j} = \frac{\mathbb{E}[\ell_t - \ell_{t+1} \mid j]}{c_{t,j} + \lambda\, d_{t,j}},$$

where $d_{t,j}$ estimates the increment to detectability contributed by revising j.

The router in Section 4 is not required to compute $v_{t,j}$ exactly; the definition specifies what the learned score s_t should approximate. Sparsity is a consequence of most regions carrying low or negative marginal value at a given state, rather than a constraint imposed for its own sake. Sparse recursion additionally supports conditional depth: covers with low channel noise can stop early, while difficult carriers consume additional steps, converting embedding computation from a fixed per-object expense into an allocation problem solved per example.

6 Training Objective

Training jointly optimizes the projection-aware decoder, the sparse recurrent encoder, and a population of steganalyzers. For sampled transformations $T_1, \dots, T_J \sim \mathcal{T}$, the payload loss is

$$\mathcal{L}_{\text{msg}} = \frac{1}{J}\sum_{j=1}^{J} \text{BCE}\big(m,\ D_\phi(T_j(y_R), k)\big).$$

The fidelity term combines a domain metric $d_{\mathcal{X}}$ with a feature-space discrepancy,

$$\mathcal{L}_{\text{fid}} = d_{\mathcal{X}}(x, y_R) + \alpha\, \|\Phi(x) - \Phi(y_R)\|_2^2.$$

Security uses an ensemble $\{A_i\}_{i=1}^{M}$ of detectors, each outputting the estimated probability that its input is stego, trained adversarially and periodically refreshed against the current encoder, following the generator-discriminator training pattern introduced for generative adversarial networks [4]. The detector objective and the encoder objective must be stated separately, since a single shared expression admits an ambiguous sign convention. Each detector is trained to maximize discrimination,

$$\mathcal{L}_{A_i} = -\mathbb{E}_{x\sim\mathcal{D}_{\mathcal{X}}} \log\big(1 - A_i(x)\big) - \mathbb{E}_{y_R} \log A_i(y_R),$$

minimized over A_i, which drives $A_i(x) \to 0$ on covers and $A_i(y_R) \to 1$ on stego objects. The encoder is trained against the current detector ensemble using the standard non-saturating confusion loss,

$$\mathcal{L}_{\text{sec}} = -\frac{1}{M}\sum_{i=1}^{M} \log\big(1 - A_i(y_R)\big),$$

minimized over θ, ϕ, ψ, which pushes $A_i(y_R)$ toward 0. This is the standard non-saturating encoder loss adapted to the present sign convention, in which the encoder's goal is $A_i(y_R) \to 0$ rather than the $D(\text{fake}) \to 1$ convention used in the original GAN formulation [4]; the two are equivalent under the relabeling $B_i = 1 - A_i$. The gradient of $\mathcal{L}_{\text{sec}}$ with respect to $A_i(y_R)$ is $1/(1-A_i(y_R))$, which is largest exactly when $A_i(y_R)$ is close to 1, the regime in which the encoder is easily detected and most needs a strong training signal. A naive alternative, minimizing $\log A_i(y_R)$ directly, also drives $A_i(y_R)$ toward 0 but has gradient $1/A_i(y_R)$ with respect to $A_i(y_R)$, which stays bounded rather than growing as $A_i(y_R) \to 1$; this is the saturating form to avoid, not $\log(1 - A_i(y_R))$, since minimizing $\log(1 - A_i(y_R))$ would push $A_i(y_R)$ toward 1 and is therefore not a variant of the same objective at all. Sparse, non-repetitive routing is encouraged by

$$\mathcal{L}_{\text{route}} = \sum_{t=0}^{R-1} \sum_{j=1}^{N} \Pr(g_{t,j} \neq 0) + \beta \sum_{t \neq t'} \frac{\langle g_t, g_{t'} \rangle}{\|g_t\|_2 \|g_{t'}\|_2 + \epsilon},$$

where g_t is the relaxed gate vector at step t, produced by a stochastic relaxation such as a hard-concrete gate for which $\Pr(g_{t,j} \neq 0)$ is available in closed form and differentiable in the gate parameters; the ℓ_0 penalty used informally elsewhere in this section is a conceptual stand-in for this expected-active-gate term, or for an ℓ_1 surrogate $\|g_t\|_1$ where a closed-form activation probability is not available, and is not itself a valid training objective, since $\|g_t\|_0$ has zero gradient almost everywhere. The second term in $\mathcal{L}_{\text{route}}$ penalizes the controller for repeatedly concentrating revisions in the same region across steps. A view-consistency term requires partial views to agree on the decoded payload distribution,

$$\mathcal{L}_{\text{view}} = \frac{1}{J(J-1)} \sum_{a \neq b} \mathrm{KL}\big(p_\phi(m \mid T_a(y_R), k) \,\|\, p_\phi(m \mid T_b(y_R), k)\big).$$

Training alternates two minimizations rather than a single min-max expression, since $\mathcal{L}_{\text{sec}}$ and $\mathcal{L}_{A_i}$ are already written with consistent signs for descent:

$$\min_{\theta,\phi,\psi} \quad \mathcal{L}_{\text{msg}} + \lambda_f \mathcal{L}_{\text{fid}} + \lambda_s \mathcal{L}_{\text{sec}} + \lambda_r \mathcal{L}_{\text{route}} + \lambda_v \mathcal{L}_{\text{view}}, \qquad \min_{A_1,\ldots,A_M} \quad \frac{1}{M} \sum_{i=1}^{M} \mathcal{L}_{A_i}.$$

The two minimizations are performed on alternating schedules, with the detector ensemble periodically refreshed and, at intervals, reinitialized in part to limit overfitting of the encoder to a fixed detector population. The hard distortion constraint of Section 2 is enforced by the projection $\Pi_{\mathcal{C}(x,B)}$ regardless of whether the learned penalties are well calibrated, which prevents the optimizer from trading visible carrier damage for payload accuracy under an imperfectly weighted objective.

7 Decoding

The receiver computes carrier features, extracts noisy projected measurements, and aggregates evidence across whatever views or regions are available. A keyed inverse module estimates codeword logits $\hat{c}$, after which the outer error-correcting decoder returns $\hat{m}$ together with a confidence score. The decoder requires no knowledge of the recursion depth or active sets used during encoding; these are properties of how the field was synthesized rather than prerequisites for reading it.

When multiple transformed copies of the stego object are available, their evidence can be aggregated before outer decoding. If the copies carry independent damage patterns, recovery can succeed even when no single copy is individually sufficient. This is a direct consequence of distributing payload evidence across both carrier coordinates and views.

8 Analysis

8.1 Graceful erasure recovery

Because S is a discrete, bounded index set, recovery probability as a function of $|S|$ cannot be continuous in the analytic sense, and bounded-noise decoding can exhibit genuine thresholds at the outer code's correction radius. The claim below is stated as monotone degradation of the best achievable decoding risk, which is what the construction actually delivers; monotonicity of the trained decoder's own error is a distinct, stronger claim addressed separately below.

Proposition 1 (Monotone degradation of achievable risk under erasure)**.** *Let O_S denote the observations available to the decoder when the surviving measurement set is S, and define the Bayes risk*

$$R(S) = \inf_{\hat{m}} \Pr[\hat{m}(O_S) \neq m],$$

the infimum taken over all decoders, not only the outer decoder actually implemented. Then $S \subseteq S'$ implies $R(S') \leq R(S)$. Let $S_\pi = \{i : U_i \leq \pi\}$ for i.i.d. uniforms U_i on $[0, 1]$, coupling the surviving sets across values of the survival probability π. Then $\pi_1 \leq \pi_2$ implies $\mathbb{E}[R(S_{\pi_2})] \leq \mathbb{E}[R(S_{\pi_1})]$, and no fixed payload block is dropped as a deterministic function of a predetermined carrier boundary, in contrast to a block-local scheme in which each payload block fails outright whenever its assigned region is erased, regardless of how much unused evidence survives elsewhere in the carrier.

Proof sketch. If $S \subseteq S'$, then O_S is a deterministic restriction of $O_{S'}$: any decoder $\hat{m}$ defined on O_S can be applied to $O_{S'}$ by first discarding the coordinates outside S, so the infimum defining $R(S')$ ranges over a superset of the effective strategies available for $R(S)$, giving $R(S') \leq R(S)$. This step uses only observation containment and holds regardless of the noise model or of Assumption 1. For the coupling, $\pi_1 \leq \pi_2$ implies $S_{\pi_1} \subseteq S_{\pi_2}$ almost surely by construction of U_i, so $R(S_{\pi_2}) \leq R(S_{\pi_1})$ pointwise on the same probability space, and taking expectations preserves the inequality. $\square$

This establishes monotonicity of the best achievable risk, not of the risk actually incurred by the trained outer decoder D_ϕ. Two further steps are needed to connect this to the deployed system, and both are empirical rather than derivable from Assumption 1 alone. First, the aggregate noise $\eta_{S'}$ does not shrink as S grows: $\|\eta_{S'}\|^2 = \|\eta_S\|^2 + \|\eta_{S' \setminus S}\|^2$ for $S \subseteq S'$, so the sufficient recovery event of Corollary 1 is not itself nested in S, even though its threshold is monotone by Assumption 1; a favorable threshold does not guarantee a favorable noise realization. Second, a fixed or mismatched decoder can in principle perform worse given additional noisy measurements, so matching the trained decoder's realized error to the Bayes-optimal monotonicity of Proposition 1 requires either an architectural guarantee, such as a decoder that provably weakly dominates its restriction to any subset of its inputs, or direct empirical verification of the kind specified in Section 9's measurement of $\gamma(S)$.

This result is conditional rather than absolute in a further sense. An adversary who removes all informative regions, destroys the carrier outright, or estimates the key-dependent projection structure can still prevent recovery. Holographic distribution changes the shape of the achievable-risk curve; it does not remove the underlying channel-capacity limit stated in Section 8.4, and it does not by itself guarantee that the trained decoder realizes the achievable risk.

8.2 Recursive dominance under uncertain response

Weak dominance of a two-step encoder over a one-step encoder follows immediately from policy-class containment: any feasible one-step policy is available to the two-step encoder as the choice of a zero second update, so the two-step optimum cannot be worse. That observation alone is close to definitional and does not explain when the dominance is strict. Strictness is a value-of-information claim, and the value of information depends on what the first step's observation actually is and what it is informative about; both must be specified precisely, since a naive reading risks overstating what a single informative observation guarantees.

Model a *persistent* carrier response as an unobserved state h, drawn from a prior $p(h)$ and unknown to the encoder before any modification is rendered, where persistence means h governs carrier susceptibility across recursion steps rather than being redrawn independently at each step. This is distinct from the identity of the transformation a future receiver will apply, which as Section 4 notes is not observable before encoding completes. The encoder's first-step observation is instead the surrogate decode

$$o_1 = \tilde{u}_1 = D_\phi^{\text{inner}}\big(\widetilde{T}_0(y_1), k\big), \qquad \widetilde{T}_0 \sim \mathcal{T},$$

which is informative about h to the extent that h and $\widetilde{T}_0(y_1)$ jointly determine the observed decode error. If h is not persistent, that is, if carrier susceptibility is redrawn independently at each step with no shared latent state, then o_1 carries no information about the state relevant to step two regardless of how informative it is about step one, and the argument below does not apply; this case is stated explicitly as a boundary of the result rather than left implicit. Let $\ell(a, h)$ be the loss of taking budget-allocation action a under persistent state h.

Proposition 2 (Dominance via value of information)**.** *Under the persistent-state model above, the two-step encoder's optimal expected loss is*

$$\mathbb{E}_{o_1}\Big[\min_a \; \mathbb{E}\big[\ell(a,h) \mid o_1\big]\Big],$$

and the one-step encoder's optimal expected loss is $\min_a \mathbb{E}[\ell(a,h)]$*. The former is no greater than the latter. Equality holds if and only if there exists a single action* $a^\star$ *that is simultaneously optimal for almost every realization of* o_1*, that is,*

$$a^\star \in \arg\min_a \mathbb{E}[\ell(a,h) \mid o_1] \quad \textit{for almost every } o_1.$$

Statistical dependence between o_1 *and* h *is necessary for strict inequality to be possible but is not sufficient: an observation can be highly informative about* h *while never changing which action is optimal, for* *instance if some fixed a_0 minimizes $\ell(a, h)$ for every h in the support of the prior. A sufficient condition for strict inequality is the existence of two events of positive probability on which o_1 induces different, uniquely optimal actions, with no single action optimal on both.*

Proof sketch. The inequality $\mathbb{E}_{o_1}[\min_a \mathbb{E}[\ell(a, h) \mid o_1]] \leq \min_a \mathbb{E}[\ell(a, h)]$ holds because the right-hand side is the loss of a single action chosen without conditioning on o_1, which is a feasible but not necessarily optimal choice for each realization of o_1 on the left-hand side; taking the expectation of the pointwise minimum cannot exceed the minimum of the expectation. Equality holds exactly when some $a^\star$ attains the conditional minimum for almost every o_1, since then the conditional policy and the unconditional policy coincide almost everywhere and their expected losses agree. Under the stated sufficient condition, no single action can attain the conditional minimum on both events, so no $a^\star$ satisfies the equality condition, and the inequality is strict. Recursion depth beyond two steps follows by the same argument applied to the posterior over h remaining after each successive observation, provided h continues to persist across the additional steps. □

This formalization makes explicit what recursion can and cannot buy. If carrier response is not persistent across steps, the surrogate decode at step t is uninformative about the state relevant to step $t + 1$, and additional recursion steps yield no improvement in expectation regardless of added computation, independent of how informative any single surrogate decode is in isolation. If carrier response is persistent but insensitive to the allocation decision in the sense that a single action is uniformly optimal, recursion likewise buys nothing. Whether realistic carrier channels supply persistent, decision-relevant response uncertainty of the kind the sufficient condition requires, and by how much, is an empirical question, addressed directly by the recursion-depth ablation in Section 9; that ablation should in particular test sensitivity to whether $\mathcal{T}$ is sampled once and held fixed across a recursion (persistent) or resampled independently at each step (not persistent), since the proposition predicts a qualitative difference between the two conditions.

8.3 Sparse modification and detectability

Sparsity does not automatically imply security. A large change concentrated in a few regions can be easier to detect than a small change spread across many. The governing quantity is the divergence between the cover and stego distributions under the realized modification policy, not the cardinality of the active set as such. Sparse routing improves the frontier only when the controller correctly identifies high-capacity regions and respects per-region amplitude constraints; the present architecture therefore treats sparsity as a learned allocation mechanism evaluated under adversarial detection, not as a security guarantee in itself.

8.4 Capacity limitation

The capacity bound below requires a specific operational model of the channel induced by $\mathcal{T}$, since $L \leq C_{\mathcal{T}}(B)$ as a statement about total bits per carrier object is dimensionally consistent but is not itself a Shannon-type asymptotic rate statement, and $\mathcal{T}$ being a declared distribution rather than a fixed map admits more than one reading. Three readings are possible, with different capacities in general: an *averaged* channel, in which both encoder and decoder know $\mathcal{T}$ and the relevant object is the capacity of the channel obtained by averaging over $T \sim \mathcal{T}$; a *compound* channel, in which some fixed but unknown T from the family is realized once and held fixed, and reliable communication must hold uniformly over the family; and an *adversarial* channel, in which T is chosen from the family by an adversary aware of the encoding strategy. Since SRHS trains against sampled transformations drawn from a declared $\mathcal{T}$, the averaged-channel reading is the one directly justified by the training objective of Section 6; the compound and adversarial readings are relevant to the active-warden and forensic-adversary cases of Section 10 and generally admit lower capacity than the averaged case.

Proposition 3 (Capacity bound)**.** *Let $C_{\mathcal{T}}(B)$ denote the capacity, in bits per carrier use, of the channel induced by transformations drawn from $\mathcal{T}$ under distortion budget B, on whichever of the three readings above is in force. For a sequence of payload lengths L_n encoded into carrier objects of size n, reliable transmission in the standard asymptotic sense requires, subject to the qualifications of channel coding [1],*

$$\limsup_{n\to\infty} \frac{L_n}{n} \le C_{\mathcal{T}}(B),$$

equivalently, any fixed communication rate R used to encode the payload must satisfy $R < C_{\mathcal{T}}(B)$ for reliable decoding to be achievable as n grows.

Projection redundancy, recursion, and learned routing cannot violate this bound under any of the three channel readings; their function is to move the achievable operating point closer to $C_{\mathcal{T}}(B)$ by using the available capacity more effectively, not to relax the bound itself. Reporting in Section 9 should state which reading of $\mathcal{T}$ a given robustness curve corresponds to, since a system evaluated only against the averaged channel has not been shown to survive the compound or adversarial case.

9 Experiments

The first evaluation should use images, since cropping, recompression, scaling, blur, noise, and color transformations provide controlled channels together with mature steganalysis baselines. Later evaluations should test audio, video, and text separately, and success in one modality must not be reported as evidence of modality-independent success.

The principal comparison holds cover distribution, payload length, key length, distortion budget, training transformations, and compute reporting constant across conditions. Baselines should include traditional transform-domain embedding, a strong one-pass learned encoder in the style of existing deep steganography systems [5, 6], a dense recurrent encoder, and a sparse non-recurrent encoder. SRHS itself must be evaluated at several payload rates and several maximum recursion depths.

Primary outcomes are bit error rate after transformation, exact-message recovery rate, detector AUC, calibrated distortion, and encoding cost. A further outcome specific to the projection stage is the empirical restricted gain $\gamma(S)$ of Assumption 1, estimated as the minimum over sampled codeword pairs and realized surviving sets S of the ratio $\|P_{k,S}(\bar{c}_a - \bar{c}_b)\|^2/\|\bar{c}_a - \bar{c}_b\|^2$, reported alongside the decoding margin threshold of Corollary 1 for the code actually in use. This directly tests whether Assumption 1 holds at the operating point used elsewhere in the evaluation, rather than being assumed to hold because the rest of the pipeline appears to work. The same measurement, repeated across nested surviving sets of increasing size, also directly tests the monotonicity claim used in Section 8.1: $\gamma(S)$ should be empirically non-decreasing as S grows under simple row restriction, and a violation would indicate that the projection is not being restricted by row deletion alone. Robustness curves should vary crop area and position, compression severity, scaling factor, and compositions of transformations. Security evaluation should include detectors seen during training, held-out architectures never seen during training, handcrafted rich-model detectors of the kind used in classical steganalysis [7], detectors trained from scratch on frozen encoder outputs, and tests conducted under key reuse; reporting results only against the adversary used in training is not sufficient evidence of security.

The decisive ablations remove one component at a time: holographic projection, outer error correction, sparse routing, recurrent residual correction, view consistency, and adversarial security training. A further control replaces adaptive routing with randomly selected regions matched for sparsity and distortion; if adaptive sparse recursion performs genuine work, it should outperform this control most clearly under heterogeneous carrier damage. A separate control, specific to Proposition 2, holds the surrogate transformation $\widetilde{T}_t$ fixed across all steps of a given recursion versus resampling it independently at each step; the proposition predicts a measurable recursion-depth benefit only in the fixed, persistent case, so equal performance across both conditions would indicate that the encoder is not in fact exploiting persistent carrier state.

Three preregistered hypotheses define success. First, at matched distortion and payload, SRHS should lower transformed-channel bit error without increasing held-out detector AUC. Second, partial-view recovery should degrade smoothly with retained evidence, consistent with Proposition 1, and should outperform localized embedding at equal redundancy. Third, recursion should improve the frontier beyond a parameter-matched one-pass model, consistent with Proposition 2, while sparse routing should reduce modification support without concentrating detectable artifacts.

Failure is equally informative and equally specified in advance. If gains disappear against detectors not used in training, the method has learned the training steganalyzers rather than steganographic security. If the projection helps only because it adds redundancy, a conventional code at matched rate is sufficient and the holographic construction is unnecessary. If additional recursion steps offer no gain after compute matching, adaptive revision is unnecessary. If robustness is obtained only by exceeding the declared distortion budget, the central claim of the paper is false.

10 Security Model and Misuse

Three observers are distinguished, extending the classical warden model of covert communication [2]. A passive warden attempts to decide whether a carrier contains a payload. An active warden applies transformations intended to destroy hidden communication while preserving ordinary utility of the carrier. A forensic adversary may collect many carriers, observe repeated key use, choose messages adaptively, or train a detector directly against a deployed encoder.

SRHS is designed primarily for robust covert communication against passive and bounded active wardens. It does not by itself provide payload confidentiality or authenticity; the payload should be encrypted and authenticated prior to outer coding and projection. Learned indistinguishability from a trained detector ensemble is likewise not cryptographic security: empirical detector failure is evidence relative to a declared adversary class, not proof against every possible test.

The same capabilities that support covert communication also support legitimate provenance marking, resilient metadata, synchronization signals, and ownership claims, and they can support malicious coordination or exfiltration. Responsible deployment should include rate controls, explicit authorization for the carriers used, auditable key management, and evaluation by independent steganalysts. Research reporting should distinguish defensive watermarking applications from covert-channel applications even where the two share underlying components.

11 Extensions

The architecture extends across carrier scales without altering its core structure. In video, measurements can be dispersed over space, time, motion trajectories, and latent scene state, while recursion revisits only frames or objects whose rendered channels prove unreliable under the internal decode. In audio, routing can follow transient masking and spectral uncertainty. In text, admissible edits are discrete and semantic constraints dominate; the projection can remain continuous internally, but realization requires a constrained generator and substantially stronger distributional testing than is needed for continuous media.

A further extension treats the carrier not as a finished object but as the output of a generative process. The payload can then influence choices among perceptually or semantically equivalent continuations, such as texture variants, camera micro-motions, mesh tessellations, lexical alternatives, or procedural seeds. The secret is encoded in a distributed pattern of otherwise admissible generative choices, and sparse recursion selects only a few decision points per pass, reevaluating recoverability after each rendering. This moves steganographic intervention upstream, from modifying a finished artifact to steering its construction.

12 Conclusion

Steganography has largely been posed as a placement problem: determine where to put a secret while changing the carrier as little as possible. Sparse Recursive Holographic Steganography poses a different problem: determine how to construct a distributed evidence field that remains recoverable across partial observations, and then synthesize that field through a small number of adaptive carrier interventions.

The architecture joins three components that are individually familiar but jointly consequential. Error-correcting projection makes payload evidence global rather than local. Sparse routing spends distortion where it carries the highest marginal value, in the sense of Section 5. Recursive correction lets the encoder observe the realized channel response and revise its allocation without increasing parameter count. None of these components removes the underlying limits of channel capacity, detectability, or adversarial adaptation stated in Section 8. Jointly, they define a credible route toward a different point on the robustness-security frontier, contingent on the experimental hypotheses of Section 9.

If those hypotheses hold under the specified ablations, the practical consequence extends beyond a new encoder architecture. Hidden information would no longer be identified with particular altered carrier locations; it would become a recoverable property of the carrier as a whole, weakly present in many views, strengthened by aggregation, and written through adaptive computation rather than fixed placement. The unit of steganographic design would then shift from the secret bit to the secret field.

A Reference Algorithm

Input: cover x, message m, key k, budget B, maximum depth R, active-set size K.

1. Encrypt and authenticate m; apply the outer error-correcting code.
2. Generate the keyed projection and compute the distributed measurements u.
3. Initialize $y_0 = x$, residual $r_0 = u$, remaining budget $b_0 = B$.
4. For $t = 0, \ldots, R-1$: score all carrier regions; select at most K into the active set; propose bounded revisions; project the result into the admissible distortion set; simulate channel transformations; internally decode the projected measurements; update the residual and budget; stop if robust confidence exceeds the threshold.
5. Return the final stego carrier y.

Decoding: extract noisy projected measurements from the available carrier view or views, apply the keyed inverse module, aggregate evidence across views, decode the outer code, and verify authentication before releasing the message.

B Minimal Reproducibility Standard

A credible report must release the exact train, validation, and test splits; the payload generation procedure; the transformation distributions; the distortion constraints; the key schedule; the encoder and detector architectures; the detector retraining protocol; the stopping policy; the random seeds; and complete robustness-detectability curves. Results must be reported at matched payload and distortion across all compared systems. Compute should be reported both per carrier and per successfully recovered payload bit. Negative results obtained under unfamiliar steganalyzers and composed transformations must not be omitted from the report.

References

[1] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 27(4):623–656, 1948.

[2] G. J. Simmons. The prisoners' problem and the subliminal channel. In D. Chaum, editor, *Advances in Cryptology: Proceedings of Crypto 83*, pages 51–67. Plenum Press, New York, 1984.

[3] E. J. Candès and T. Tao. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, 2006.

[4] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, volume 27, pages 2672–2680, 2014.

[5] S. Baluja. Hiding images in plain sight: deep steganography. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

[6] J. Zhu, R. Kaplan, J. Johnson, and L. Fei-Fei. HiDDeN: hiding data with deep networks. In *Computer Vision – ECCV 2018*, Lecture Notes in Computer Science, volume 11219, pages 682–697. Springer, Cham, 2018.

[7] J. Fridrich and J. Kodovský. Rich models for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, 7(3):868–882, 2012.