EPHEMERAL SPHERICAL MODULATION: WATERMARKING WHERE THE DIFFUSION CHANNEL CANNOT LOOK

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion watermarks written into the initial noise must keep it standard Gaussian while the message survives generation, image processing and inversion, and existing constructions are compared pipeline by pipeline under assumed noise models. We ask what the pipeline observes: under ideal secret frames, the same Gaussian latent for every prior-preserving watermark, so watermarks differ only in how their keys read one channel. The prior then reduces to a Gram invariant of the frame-sharing pattern. Without a noise model, the decoder reads the channel through a uniformly random consistent frame, and on acutely recovered blocks equal magnitudes minimize expected sign errors among profiles with no coordinate above a quarter of the block energy. Equal magnitudes and an exact prior force a separate frame for every block of every image: ephemeral spherical modulation (ESM), whose frames are regenerated from a key and per-image nonce, with half-normal modulation as its width-one case. On SD1.5 and SD2.1-base, ESM recovers 114/128 and 110/128 complete 512-bit messages under JPEG50, versus 69/128 and 74/128 for half-normal modulation. Channel samples from one watermark's images reproduce another's recovery rates without fitted parameters, and the tested crop leaves recovered blocks nearly orthogonal to their sources, putting every readout of the family at chance. These results concern payload recovery with the correct key and nonce, not calibrated detection, synchronization or cryptographic security.

> *"Mathematicians do not study objects, but the relations between objects; to them it is a matter of indifference if these objects are replaced by others, provided that the relations do not change."*—Henri Poincaré, *Science and Hypothesis*, 1905 (explained in Appendix I)

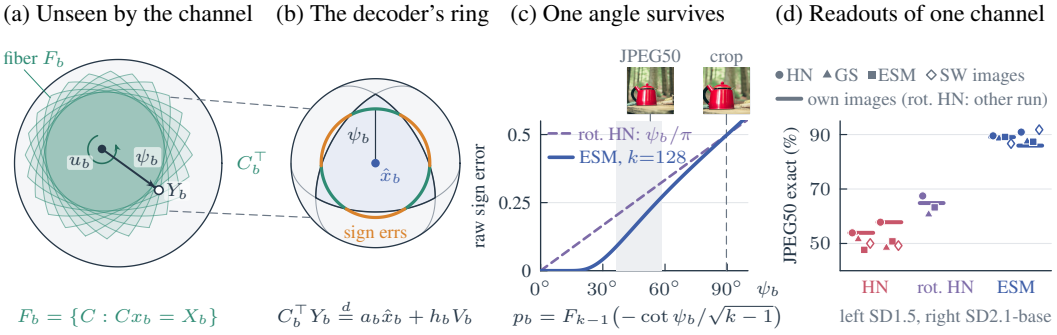


Figure 1: **Under ideal secret frames, every prior-preserving watermark reads one channel through the fiber of its key.** (a) The frames consistent with X_b spin the sign cell about u_b, unseen by the channel (Proposition 2; $k = 3$). (b) The decoder sees Y_b uniformly on a ring at angle $\psi_b = 40°$; signs err on the orange arcs. (c) Laws at $k = 128$ (Theorems 3 and 4(a)); band: JPEG50 interquartile angles; dashed: crop median; images: SD1.5 outputs. (d) JPEG50 exact recovery on each construction's channel samples (markers; SW descriptive) and on own images (bars), with no fitted parameter (Section 5.1).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

1 INTRODUCTION

Diffusion models turn Gaussian noise into images, and a watermark can be written into that noise before generation begins (Wen et al., 2023; Yang et al., 2024a; Hu et al., 2026). The generator is unchanged; the decoder inverts the image to an estimate of the noise and reads the message there. Two requirements pull against each other. The noise must remain standard Gaussian, the distribution the generator was trained on, and the message must survive sampling, compression, resizing and the approximate inversion. For a 512-bit identifier the second requirement is severe: an estimate that is correct on 99.8% of the bits still fails whenever one bit is wrong.

Existing constructions combine signs, magnitudes and rotations differently. Gaussian Shading keeps Gaussian magnitudes and randomizes signs with a key (Yang et al., 2024a); Spherical Watermark gives all coordinates of a block one magnitude and rotates every block with one secret orthogonal frame (Hu et al., 2026). Each is evaluated by running it through an image pipeline and explained, if at all, with an additive-noise model that the pipeline does not obey. A measured gain may therefore belong to the watermark, or to the way one pipeline distorts one kind of latent.

We begin with a naive question: what does the pipeline actually see? The generator never sees the key. If the secret frames are ideal random rotations, the latent it receives is exactly standard Gaussian for every message and every prior-preserving watermark, and the joint law of the transmitted and recovered latents is the same for all of them. In this ideal model the pipeline cannot tell these watermarks apart, nor any of them from unwatermarked noise. What distinguishes them lies where the pipeline cannot look: in the set of frames that remain consistent with the transmitted latent, the *fiber* over the latent. A watermark is a way of reading one common channel through that fiber.

This viewpoint turns three design questions into statements about the fiber, none of which needs a model of the channel. *Does the latent keep its prior?* Exactly when the Gram matrix of the blocks that share a frame has its Gaussian law; sharing within an image, across images or through a reused nonce exposes inner products that an equal-magnitude code places on a lattice (Proposition 1). *What does image processing do to the message?* The decoder sees the recovered latent through a frame drawn uniformly from the fiber, so for equal-magnitude rotations of width k the error law of each coordinate depends only on the recovered block's angle to its source (Proposition 2, Theorem 3). *Which code is best?* Independent frames replace the channel by its average over block orientations, and on the averaged channel equal magnitudes maximize the coherent soft sum and minimize expected sign errors at acutely recovered blocks, among codes that spread each block's energy (Theorem 4).

These answers leave little freedom in the frames. The decoder wants equal magnitudes, and once magnitudes are equal, an exactly Gaussian latent forces one frame for every block of every image and chi-distributed radii (Proposition 1(iii)); the block width and the repetition code remain design choices. Ephemeral spherical modulation (ESM) is this construction made practical: it regenerates the frames from a shared key and a public per-image nonce through Householder factors, without stored matrices or cubic-cost QR. The natural control, half-normal modulation (HN), is the same construction with blocks of width one. Its fiber is a point, so it reads the channel coordinate by coordinate, and under Gaussian noise it has the Cauchy tail of the width-one member (Corollary 5).

On two Stable Diffusion backbones with a paired 512-bit protocol, ESM recovers 114/128 and 110/128 complete messages under JPEG50, against 69/128 and 74/128 for HN. Because the channel is common, each decoder can be evaluated on channel samples measured with another watermark's images, and these transplants reproduce the decoders' own recovery rates within sampling error, without fitted parameters (Section 5.1). Widening HN's frames, a step without a channel-free sign, helps under JPEG50 mostly as spatial interleaving, and equalizing magnitudes in fixed frames adds 24.2 points of exact recovery. The angular law reproduces ESM's exact recovery from each image's measured geometry within about 3 points in ten settings, and locates the one channel on which every method fails: after the tested crop, recovered blocks are nearly orthogonal to their sources, so every readout of the family is at chance.

Contributions.

1. **A fiber view of initial-noise watermarking.** A Gram-invariant criterion for exact prior preservation covering sign masks, shared and per-image frames and nonce reuse, and a fiber principle that gives an angular recovery law for every magnitude profile without a noise model.

2. **Channel-free design and its construction.** Symmetrization by independent frames, which turns HN into random-hyperplane rounding, and optimality of equal magnitudes among spread profiles on acutely recovered blocks, with a counterexample marking its limit; ESM realizes both with $O(Bk^2)$ frame regeneration from a key and a nonce.
3. **Evidence at the level of the channel.** Complete-message recovery on two backbones, a parameter-free channel transplant across watermarks, a two-step ablation separating symmetrization from equalization, and angular-law reproductions of recovery that locate the failure under cropping.

2 Related work

Diffusion watermarking. Learned encoders embed messages in finished images (Zhu et al., 2018; Tancik et al., 2020), and Stable Signature fine-tunes the latent decoder (Fernandez et al., 2023). Initial-noise methods leave the generator fixed: Tree-Ring writes Fourier patterns (Wen et al., 2023), RingID extends it to multi-key identification (Ci et al., 2024), Gaussian Shading maps payload bits to Gaussian samples (Yang et al., 2024a), and Gaussian Shading++ addresses nonce handling in deployment (Yang et al., 2025). Spherical Watermark combines binary angular codes, one secret rotation and random radii (Hu et al., 2026); correlation-aware angular watermarking studies dependence in latent angular codes (Zheng et al., 2026). In the terms of Section 3, these methods differ in their code and frame pattern, and ESM is the finest frame pattern with an equal-magnitude code.

Invariance and random rotations. That an invariant law is determined by the law of a maximal invariant is classical in statistics (Eaton, 1989, Section 4.1); Proposition 1 applies it to the groups generated by frame sharing. Haar frames are sampled by the subgroup algorithm (Mezzadri, 2007). Rotating a multidimensional constellation so that any two points differ in every coordinate, with the coordinates interleaved so that they fade independently, yields diversity against fading (Boutros & Viterbo, 1998); Theorem 4(a) describes the analogous averaging produced by secret random block frames. Averaging many watermarked images can expose content-agnostic patterns (Yang et al., 2024b), and fourth-order and independent-component methods can learn shared frames (Hyvärinen, 1999; Zhai et al., 2020); Proposition 1 identifies what such methods can exploit.

WAVES separates detection from identification across image changes (An et al., 2024); we measure complete-payload recovery. Appendix H gives further background.

3 Watermarks as readouts of one Gaussian channel

3.1 Codes, frames and the channel

We view the diffusion latent as N blocks of width k, $X = (X_1, \dots, X_N)$ with $X_b \in \mathbb{R}^k$; one image in our main setting has $N = B = 128$ blocks of width $k = 128$. An initial-noise watermark has two ingredients. Its *code* maps a message m to a random pre-rotation array $x = (x_1, \dots, x_N)$, which fixes the signs and magnitudes of the coordinates. Its *frame pattern* is a partition π of the blocks: blocks in the same part P share a secret orthogonal frame $C_P \in O(k)$, and the transmitted latent is $X_b = C_P x_b$ for $b \in P$. Everything between X and the decoder's estimate—generation, image processing and inversion—forms the *channel* $Y = \Phi(X, \varepsilon)$, where ε collects its own randomness, such as the prompt and the sampler. We analyze two assumptions.

(A1) Ideal frames. Frames of different parts are independent and Haar-distributed on the full orthogonal group $O(k)$, independent of the code. **(A2) Frame-blind channel.** The channel randomness ε is independent of the message, the code's randomness and the frames, jointly.

A code is *prior-preserving* if $X \sim N(0, I)$ for every message under (A1). Assumption (A2) holds by construction when prompts and sampler seeds are chosen without reference to the watermark, because the generator never sees the key. Assumption (A1) idealizes the keyed pseudorandom frames of Section 3.3; statements about a fixed finite key require a separate security argument. Table 1 places the constructions we compare in this design space.

Table 1: **The design space.** A construction is a code (magnitudes) with a frame pattern; the last two columns anticipate Proposition 1. Width-one frames are sign masks, drawn per image as in our experiments; Spherical Watermark also mixes the payload with padding bits.

Construction	k	Frame shared by	Magnitudes	Visible invariant	Exact prior
HN, Gaussian Shading	1	one coordinate of one image	half-normal	$\lvert X_j \rvert$	yes
Spherical Watermark	128	all blocks and images of a key	equal	all inner products	no (lattice)
Rotated HN	128	one block of one image	half-normal	block norms	yes
ESM	128	one block of one image	equal	block norms	yes
ESM, reused nonce	128	one block of two images	equal	cross-image products	no (lattice)

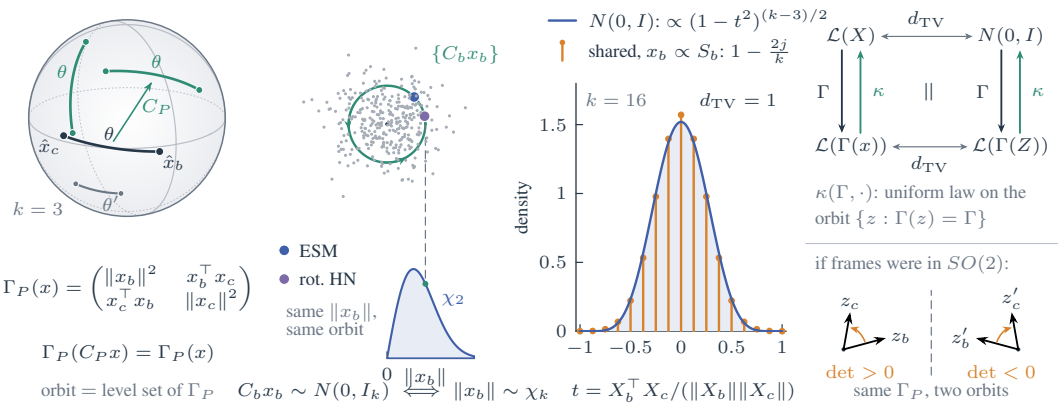


Figure 2: **Under (A1), the latent's law depends on the code only through a Gram matrix (Proposition 1).** (a) A shared frame moves blocks rigidly: the orbit of x is the level set of Γ_P, all pairs at angle θ ($k = 3$). (b) One frame per block leaves only norms visible; independent χ_k radii give an exact Gaussian (part (i); $k = 2$). (c) Equal magnitudes under a shared frame put $t = 1 - 2H(S_b, S_c)/k$ on a lattice (Theorem 7), matching $N(0, I)$'s moments yet $d_{\mathrm{TV}} = 1$ (part (ii); $k = 16$). (d) One kernel κ, uniform on the level set, proves both directions.

3.2 What the latent reveals

Averaging a point over a compact group of rotations produces the uniform distribution on its orbit. Shared frames act on the blocks of a part simultaneously, and the orbits of that action are the level sets of the blocks' inner products. The prior question therefore reduces to a single invariant.

Proposition 1 (The Gram invariant)**.** *Assume (A1) and fix a message. For each part P let $\Gamma_P(x) = (x_b^\top x_c)_{b,c \in P}$, and let $Z \sim N(0, I)$. The law of X depends on the code only through the joint law of $\Gamma(x) = (\Gamma_P(x))_{P \in \pi}$, and*

$$d_{\mathrm{TV}}\big(\mathcal{L}(X),\, N(0, I)\big) = d_{\mathrm{TV}}\big(\mathcal{L}(\Gamma(x)),\, \mathcal{L}(\Gamma(Z))\big).$$

Thus $X \sim N(0, I)$ exactly when the matrices $\Gamma_P(x)$ are independent across parts, each distributed as the Gram matrix of $|P|$ independent $N(0, I_k)$ vectors. In particular: (i) with one frame per block, the prior holds if and only if the block norms $\|x_b\|$ are independent χ_k variables, as for HN ($k = 1$), rotated HN and ESM; (ii) if $k \geq 2$, two blocks $b \neq c$ share a frame and the code has equal magnitudes within blocks, $x_b = r_b S_b / \sqrt{k}$ with $r_b > 0$ and $S_b \in \{-1, 1\}^k$, then $X_b^\top X_c / (\|X_b\| \|X_c\|) = S_b^\top S_c / k$ lies in the lattice $\{1 - 2j/k\}_{j=0}^k$, and the law of X is singular with respect to $N(0, I)$; (iii) hence, for $k \geq 2$, an equal-magnitude code yields an exactly Gaussian latent for every message if and only if every block has its own frame and, for every message, the radii r_b are independent χ_k variables.

The proof (Appendix B) is the orbit observation followed by data processing in both directions, and it settles the prior question for every row of Table 1 at once. One frame per block leaves only block norms visible; any sharing exposes inner products, which equal magnitudes place on a lattice

(Figure 2): a key-fixed frame as in Spherical Watermark, a frame per image, or a reused nonce. Since the two distances are equal, no test on the latent has smaller total error against a Haar key than the best test on its Gram invariants; the Gram-grid detector of Appendix F tests this invariant directly.

3.3 Ephemeral spherical modulation

ESM takes the finest frame pattern, one frame per block and per image, and the code that Section 4 identifies as best on it. Let $m \in \{0,1\}^p$ be the message. We map its bits to signs, repeat the sign vector r times and divide the resulting $d = pr$ coordinates into B blocks of width k, so that each copy of the payload spans $q = p/k$ blocks, $B = qr$, and the repetitions of a bit lie in distinct blocks. The main setting uses $p = 512$, $r = 32$ and $B = k = 128$, hence $q = 4$, $d = 16384$ and a payload rate of $1/32$ bits per latent coordinate. Write $S_b \in \{-1,1\}^k$ for the signs of block b.

Code and frames. Draw $G \sim N(0, I_d)$ and put $R_b = \|G_b\|$, so $R_b^2 \sim \chi_k^2$. The encoded block is

$$X_b = \frac{R_b}{\sqrt{k}}\, C_{K,n,b} S_b, \tag{1}$$

so every coordinate of a block starts at the same distance $R_b/\sqrt{k}$ from its sign boundary: $S_b/\sqrt{k}$ is the center of its sign cell, the unit vector with these signs whose smallest coordinate is largest. Proposition 1(i) gives an exact Gaussian latent under (A1). The frame $C_{K,n,b}$ is determined by a shared 32-byte key K, a unique public 12-byte nonce n per image, and the block index. We build it with the classical subgroup construction (Mezzadri, 2007): a product of k Householder reflections, each generated from a domain-separated ChaCha20 stream. Encoding applies the reflections and decoding applies them in reverse order. Generation and application cost $O(Bk^2)$ arithmetic, compared with $O(Bk^3)$ for regenerating dense QR frames; both use $O(Bk^2)$ temporary storage, and no frame is stored between encoding and decoding.

Decoding. The decoder inverts the processed image with a fixed empty-prompt pipeline to obtain Y, regenerates the frames and computes $U_b = C_{K,n,b}^\top Y_b$. The soft decoder sets $\widehat{m}_j = \mathbf{1}\{\sum_{i \in I_j} U_i > 0\}$ over the r coordinates I_j carrying bit j, with ties decoded as zero; the hard decoder counts positive signs instead. Algorithm 1 and Appendix A give the coordinate order, sampling conventions and nonce accounting; our experiments supply the nonce externally.

The controls are members of the family. Half-normal modulation (HN) sets $X_j = M_j S_j |G_j|$ with a pseudorandom sign mask M. Since $O(1) = \{-1,1\}$ and $|G_j| \sim \chi_1$, this is exactly (1) with blocks of width one. Rotated HN keeps HN's magnitudes but uses ESM's frames, $X_b = C_{K,n,b}(S_b \odot |G_b|)$. The three share the payload, the Gaussian draw G and hence every block energy, and the soft decoder; rotated HN and ESM also share their frames. From HN to rotated HN the frames widen; from rotated HN to ESM the magnitudes equalize.

4 What the decoder reads

4.1 The fiber of the key

Under (A1), a transmitted block does not determine its frame. Given the codeword x_b, the frames consistent with X_b form the *fiber* $F_b = \{C \in O(k) : Cx_b = X_b\}$, a coset of the stabilizer of x_b and a copy of $O(k-1)$; for $k = 1$ it is a single sign. The next statement makes the question of the introduction precise.

Proposition 2 (Fiber principle). *Assume (A1)–(A2) with one frame per block. Then: (i) if the code is prior-preserving, the joint law of (X, Y) is the same for every message and every such code; (ii) conditional on (x, X, Y), the frames are independent and each is uniform on its fiber F_b; (iii) with $u_b = X_b/\|X_b\|$, $a_b = u_b^\top Y_b$, $h_b = \|Y_b - a_b u_b\|$ and $\hat{x}_b = x_b/\|x_b\|$, conditional on (x, X, Y),*

$$C_b^\top Y_b \overset{d}{=} a_b\, \hat{x}_b + h_b\, V_b, \qquad V_b \sim \mathrm{Unif}\big(\mathbb{S}^{k-1} \cap \hat{x}_b^\perp\big),$$

independently across blocks.

Part (i) is immediate from Proposition 1; its use is to license comparing codes on identical channel samples (Section 5.1). Everything that separates two prior-preserving watermarks happens after the channel, in parts (ii)–(iii), which do not need prior preservation. The decoder sees the recovered block

through a uniformly random frame on the fiber: the component a_b along the transmitted direction is kept, and the perpendicular component of length h_b is spread uniformly over the directions orthogonal to the codeword. For HN the fiber is a point, $h_b = 0$, and the decoder reads each coordinate's own agreement $\operatorname{sign}(X_j)Y_j$.

4.2 THE ANGULAR LAW

Theorem 3 (Angular decoding law). *Under the assumptions of Proposition 2, let $k \geq 3$ and write each codeword as $x_b = \|x_b\|\,(S_b \odot \alpha_b)$, with payload signs $S_b \in \{-1, 1\}^k$ and a magnitude profile $\alpha_b \geq 0$, $\|\alpha_b\| = 1$, $\alpha_{bj} < 1$. Let F_d be the CDF of one coordinate of a uniform point on $\mathbb{S}^{d-1}$, and let $\psi_b \in [0, \pi]$ be the recovered angle, $\cot \psi_b = a_b/h_b$. Conditional on (x, X, Y) with $h_b > 0$, the decoded sign of coordinate j of block b differs from S_{bj} with probability*

$$p_{bj} = F_{k-1}\left(-\cot\psi_b \frac{\alpha_{bj}}{\sqrt{1-\alpha_{bj}^2}}\right), \tag{2}$$

errors in distinct blocks are independent, and the error count of a bit repeated in distinct blocks is Poisson-binomial. With equal magnitudes, $\alpha_{bj} = 1/\sqrt{k}$ and $p_b = F_{k-1}\big(-\cot\psi_b/\sqrt{k-1}\big)$ depends on the block only through ψ_b; moreover the aligned block $\operatorname{diag}(S_b)C_b^\top Y_b$ equals $(a_b/\sqrt{k})\mathbf{1} + h_b W_b$ in law, with W_b uniform on the unit sphere of $\mathbf{1}^\perp$.

The last identity gives the joint law of all soft scores, including the dependence between bits that share a block, and hence the probability of recovering a complete message (Corollary 6). The pipeline supplies the pair (X, Y) and the fiber supplies the law of the decoded bits; no model of the channel enters. We call the equal-magnitude case the cap law. An orthogonal block makes each sign a fair coin for every profile.

4.3 DESIGN ON THE FIBER

The fiber principle turns design choices into statements that hold for every frame-blind channel at once. We call a magnitude profile *spread* if no coordinate carries more than a quarter of the block energy, $\max_j \alpha_{bj}^2 \leq 1/4$.

Theorem 4 (Design on the fiber). *Assume (A1)–(A2), with a separate frame or sign mask for every block. (a)* Symmetrization. *For any channel Φ, the decoded bits of rotated HN have the joint law of those of HN under the channel averaged over independent Haar rotations O of the blocks, $X \mapsto O^\top \Phi(OX, \varepsilon)$. Conditional on (X, Y), each raw sign of rotated HN in block b errs with probability ψ_b/π. (b)* Coherent gain. *Conditional on (x, X, Y), the aligned block $\operatorname{diag}(S_b)C_b^\top Y_b$ has mean $a_b\alpha_b$ and total variance h_b^2, whatever the magnitudes. If $a_b > 0$, its mean averaged over the k coordinates is at most $a_b/\sqrt{k}$, with equality exactly for equal magnitudes. (c)* Equal margins. *Let $k \geq 5$ and $a_b, h_b > 0$. Among spread profiles, equal magnitudes minimize the conditional expected number of sign errors in the block; the inequality reverses when $a_b < 0$. The restriction to spread profiles cannot be dropped: as $a_b/h_b \to 0^+$, a one-hot profile has fewer expected errors.*

Part (a) is what widening the frames buys: the decoder no longer sees how the channel's distortion is oriented within blocks, only its average over orientations, which can be worse than a particular orientation; a positive rescaling of each coordinate preserves every HN sign but not the rotated ones. Rotated HN is then random-hyperplane rounding (Goemans & Williamson, 1995): each coordinate compares X_b and Y_b along an independent uniformly random direction and errs exactly when that hyperplane separates them. Parts (b)–(c) are what equal magnitudes buy on the symmetrized channel: the largest coherent soft sum at unchanged total noise, the same conditional law for every bit of a repetition group, and, since $v \mapsto F_{k-1}(-c\sqrt{v/(1-v)})$ is convex on $[0, 1/4]$ for every $c > 0$, the fewest expected raw sign errors. ESM replaces the hyperplane by the cap law, which at $k = 128$ lies below ψ_b/π at every acute angle (numerically; part (c) gives it up to a factor $1 + 1.6 \times 10^{-9}$): at $47°$, near the median JPEG50 angle, a raw sign errs with probability 0.18 against 0.26. Both codes share the law of recovered angles (Proposition 2(i)), so ESM makes fewer expected raw sign errors than rotated HN on every frame-blind channel with acute recovered blocks. The passage HN $\to$ rotated HN $\to$ ESM is therefore one step whose sign depends on the channel followed by one whose sign, for

raw sign errors, does not (Section 5.2). Under additive Gaussian noise the family is ranked explicitly (Corollary 5, Appendix B): width k errs with probability $\Pr\{t_k < -1/\sigma\}$, and HN, the case $k = 1$, has the heaviest tail, $\arctan(\sigma)/\pi$.

5 EXPERIMENTS

Protocol. We embed 512-bit messages on SD1.5 and SD2.1-base, crossing eight test keys with sixteen authored prompts for 128 paired requests per model and method. Generation uses 50 DPM-Solver steps at guidance 7.5; inversion uses the VAE posterior mode and 50 empty-prompt DDIM steps at guidance 1 (Lu et al., 2022a;b; Song et al., 2021). The channels are the clean image, JPEG at quality 50 and 30, resizing to 256^2 and back, and a centered 410^2 crop resized to 512^2. Bit accuracy (BA) is the fraction of correct bits, and exact recovery requires all 512. HN and ESM share payloads and Gaussian draws and use soft sums; Gaussian Shading (GS) and Spherical Watermark (SW) keep source-based adapters and their native hard decoders (Appendix A). The prespecified primary comparison is the ESM–HN BA difference under JPEG50. Intervals are pointwise 95% crossed-bootstrap intervals over keys and prompts with 5000 draws, retaining method pairing (Owen, 2007).

5.1 ONE CHANNEL, MANY READOUTS

Proposition 2(i) implies that a stored pair (X, Y) from one construction's image is a valid channel sample for every prior-preserving construction. We therefore evaluate each decoder on pairs measured with other constructions' images. For HN the fiber is a point, so its decoder is a deterministic function of the pair; for rotated HN and ESM we average over the fiber with Proposition 2(iii), using 512 joint draws per pair. No parameter is fitted, and the analysis plan was fixed before any transplanted outcome was computed (Appendix C.1). The HN readout reproduces all 1280 logged HN outcomes on HN's own pairs.

Figure 1(d) and Table 5 show the result under JPEG50. Across sources, each readout gives similar rates: ESM's readout recovers 88.6–89.5% of messages on SD1.5 images of HN, GS and ESM (observed 89.1%) and 87.4–90.9% on SD2.1-base (observed 85.9%), and rotated HN, whose images were never used, is predicted at 60.7–67.5% against 64.8% observed in a separate execution. HN's transplanted rates, 47.7–51.6% and 48.4–50.8%, lie below its observed 53.9% and 57.8%; these differences share one observed reference and are of the size of the 69-versus-64 variation between two executions on identical source latents. Over both backbones and all channels, 44 of the 46 cross-construction intervals for exact recovery contain zero, the exceptions being two ceiling cells (128/128 observed, 99.2% and 99.8% predicted), and bit-accuracy differences are at most 0.49 points. Spherical Watermark sources, outside the premise, agree as well. Readouts of identical channel samples, a descriptive comparison outside the fixed plan that removes image-to-image variation, differ by 33–41 points between HN and ESM on Gaussian-latent sources, with HN $<$ rotated HN $<$ ESM wherever all three are available.

5.2 WIDEN THE FRAMES, THEN EQUALIZE THE MAGNITUDES

A separate SD1.5 execution regenerates HN, rotated HN and ESM together on the 128 paired requests (Table 2). Widening HN's frames to $k = 128$ changes JPEG50 exact recovery by $+14.8$ points with an interval that includes zero; Theorem 4(a) gives this step no channel-free sign, and on identical channel samples (Table 5) it is favorable under JPEG50, by 9–16 points. Most of it is spatial interleaving: HN puts all copies of a bit in one latent column, and a random interleaver gains 7–14 of these points on the same samples, leaving 1.5–1.9 for rotation; ESM still beats interleaved HN by 24–30 points on SD1.5 and 19–25 on SD2.1-base, all intervals excluding zero (Appendix C.3). Equalizing the magnitudes in the same frames adds 24.2 points, with interval $[10.9, 38.3]$, and the ESM advantage is positive in every key group. Theorem 4(c) gives this step a channel-free sign for expected raw sign errors on acute blocks and part (b) a larger coherent soft sum; the complete-message gain is consistent with both. The premise holds in the main study: under JPEG50, 99.3–99.8% of recovered blocks are acute (Table 8), and the rest lie within 13° of a right angle, where the reversal costs at most 0.02 per coordinate. The fourth corner of this design, equal magnitudes behind sign masks, is excluded by Proposition 1(i).

Table 2: **Widen, then equalize.** A separate SD1.5 execution on the 128 paired requests; each row changes one ingredient of the row above. Cells give BA (%) / exact messages; the last column is the JPEG50 exact-recovery change from the row above, in points, with its 95% interval.

Code	Change	Clean	JPEG50	Δ exact [95% CI]
HN	width-1 frames	99.98 / 123	99.19 / 64	
Rotated HN	widen frames to $k = 128$	99.99 / 125	99.50 / 83	$+14.84$ $[-2.34, +34.38]$
ESM	equalize magnitudes	100.00 / 128	99.81 / 114	$+24.22$ $[+10.94, +38.28]$

5.3 The angular law and where it fails

Theorem 3 and Corollary 6 turn each ESM source/recovered pair into a conditional probability of recovering the complete message, without channel parameters. Over the 1280 ESM inversions, predicted and observed JPEG50 exact rates are 89.04%/89.06% on SD1.5 and 87.15%/85.94% on SD2.1-base, and the largest gap across the ten model–channel settings is 3.04 points (Figure 3); an independent re-simulation with other seeds gives 3.22, and the Monte Carlo standard error of each rate is below 0.2 points. The same geometry locates the one channel on which every method fails. After the crop, the median recovered block is within half a degree of orthogonal to its source and 47–48% of blocks are obtuse, so every coordinate is close to a fair coin for every magnitude profile and no code of this family can recover the message; the failure is one of synchronization, not of coding.

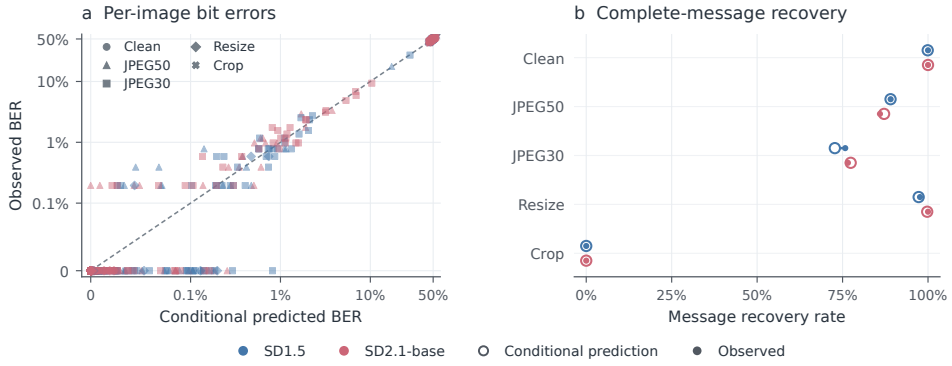


Figure 3: **The angular law tracks ESM recovery.** (a) Observed versus predicted soft BER for 1280 image/channel pairs. Colors distinguish models and shapes distinguish channels; axes are linear below 0.1% and logarithmic above. (b) Observed exact-recovery rates (filled) and mean conditional probabilities (open), with 128 inversions per setting. Predictions use 512 joint draws conditioned on the source and recovered latents, preserving within-block dependence without fitting channel parameters.

5.4 Comparison with Gaussian Shading and Spherical Watermark

Table 3 compares the four constructions in the common pipeline. Under JPEG50, ESM recovers 114/128 messages on SD1.5 and 110/128 on SD2.1-base, versus 69/128 and 74/128 for HN, paired advantages of 35.16 points ($[17.97, 52.34]$) and 28.12 points ($[11.72, 46.88]$). The prespecified BA differences are only 0.63 and 0.65 points, so small bit-level gains produce many more intact messages. ESM also exceeds the GS and SW adapters, including GS with soft decoding (68 and 79 exact recoveries). Under SW's own numerical settings, full-precision SD1.5 with sampled VAE inversion, ESM recovers 115/128 JPEG50 messages against SW's 32/128 (Appendix C.6). All four constructions recover no complete message after the crop. Because every method uses repetition only, we also evaluate an idealized Reed–Solomon outer code on the measured error patterns: it narrows but does not close the gap, and at 480 net bits ESM recovers 96.9% and 93.0% of JPEG50 messages against 69.5% and 73.4% for HN (Appendix C.2).

Table 3: **512-bit recovery across image transformations.** Each cell reports BA (%) / exact messages out of 128. Color encodes exact recovery on a common scale: blue at 0%, white at 50%, rose at 100%. Bold and underlined counts mark the best and second-best recovery within each model and channel; all-zero columns are unranked.

Method	Clean	JPEG50	JPEG30	Resize	Crop
SD1.5					
HN	99.98 / 124	99.19 / 69	98.27 / 49	99.88 / 106	51.16 / 0
GS	99.96 / 117	98.66 / 46	97.39 / 34	99.71 / 85	50.64 / 0
SW	99.67 / 111	93.60 / 35	88.83 / 19	98.12 / 64	49.97 / 0
ESM	100.00 / **128**	99.82 / **114**	99.60 / **97**	99.99 / **125**	51.29 / 0
SD2.1-base					
HN	99.98 / 125	99.23 / 74	98.50 / 50	99.91 / 109	50.76 / 0
GS	99.98 / 119	98.89 / 45	97.54 / 33	99.84 / 85	50.82 / 0
SW	99.91 / 115	94.23 / 35	89.30 / 22	99.22 / 68	50.16 / 0
ESM	100.00 / **128**	99.87 / **110**	99.57 / **98**	100.00 / **128**	51.49 / 0

5.5 What the latent reveals, image quality and cost

As Proposition 1 predicts, on the implemented source sampler and over eight held-out keys, a statistical-feature classifier and a Gram-grid detector separate a synthetic shared-frame construction from Gaussian noise with AUC 1.000, while ESM gives 0.480 $[0.460, 0.501]$ and 0.525 (Appendix Table 18); a within-key fourth-order learner failed its sensitivity control (Appendix C.8).

ESM's CLIP text alignment differs from Gaussian noise by -0.0006 and $+0.0006$ on the two backbones, with paired intervals that include zero; its empirical FID is close to the distance between two unwatermarked Gaussian draws, and paired KID and CLIP-MMD intervals include zero. A CLIP-feature classifier does separate ESM from Gaussian images on SD2.1-base (AUC 0.582 $[0.551, 0.622]$, against 0.489 for two Gaussian draws; Appendix C.9). Under diffusion regeneration at strength 0.1, ESM keeps 24/32 complete messages against HN's 16/32 (interval $[-12.5, 62.5]$), and both nearly fail at 0.3 (Appendix C.10). Householder encoding and decoding take 49.8 and 49.3 ms, 3.8–3.9 times faster than dense QR regeneration and about 50 ms more than HN (Appendix C.11).

6 Discussion and conclusion

Under ideal frames a diffusion pipeline sees only a Gaussian, so the message lives in the fiber of frames consistent with the latent and is read through an angle; ESM's recovery, transplants and failures behave as that fiber predicts. The prior becomes a condition on the Gram invariant of shared frames, recovery a readout governed by the recovered angle, and design a choice of how large the fiber is and how the message sits in it. The two halves of the design are tied: the flatness that equalizes margins is also what a shared frame exposes, through the Gram lattice and the fourth-order contrast of Appendix F (Theorem 8), which is why equal magnitudes admit no frame sharing at all.

Limitations. The distributional statements assume ideal Haar frames; hiding under a fixed finite key needs a separate security analysis. The angular law explains observed source/recovered pairs, but predicting a new channel's geometry from images, and optimizing codes for complete messages, remain open. The evaluation uses two backbones and authored prompts; recovery fails under the tested crop and drops under stronger regeneration, so synchronization and removal attacks remain open. The GS and SW adapters are not the complete published systems, the outer-code comparison uses measured error patterns rather than embedded codewords, and intervals exclude variation between executions (HN recovered 69 and 64 of 128 JPEG50 messages in two executions). Decoding assumes the correct key and an external nonce, whose reuse is singular for ESM (Proposition 1); attribution also needs authenticated nonce transport, false-positive calibration and resistance to forgery.

ETHICS STATEMENT

Watermark recovery alone does not establish ownership, and a missing watermark does not establish that an image is human-generated. Deployment requires calibrated error rates, authenticated key/nonce handling and safeguards against false attribution. Our experiments use authored prompts and model-generated images with the generation safety checker enabled. Image panels illustrate the evaluated transformations and make no claims about depicted real people.

REPRODUCIBILITY STATEMENT

The appendices specify the model sources, generation and inversion settings, decoding rules, data splits and statistical analysis. The accompanying code contains the implementations, evaluation scripts and detailed environment configuration, together with the channel-transplant re-analysis and the protocol fixed before it was run. Development and test keys and prompts are disjoint.

AI USE STATEMENT

The research question, the choice of experiments and the decisions taken at every stage of this work are the authors'. We used LLM-based assistants throughout as tools, including for proposing the reformulation of Sections 3–4 in terms of Gram invariants and fibers, for drafting the corresponding statements and proofs, and for designing and implementing the channel-transplant re-analysis; for coding assistance in implementing the experiment and analysis scripts and the figures, for drafting and editing text to improve readability, for formatting references, for suggesting titles and keywords, and for assistance in developing and checking the theoretical arguments, including drafting proof steps that the authors then verified.

We have reviewed all AI-assisted work: every theoretical claim was checked by hand and, where it admits a numerical test, against the numerical diagnostics in Section 5 and Appendix C.16; every reported experimental number is regenerated from frozen campaign records by the accompanying analysis pipeline; and the literature discussion was checked against the cited sources. We take responsibility for the final content of this work, including text, claims, code and artifacts produced with the aid of generative AI.

REFERENCES

Bang An, Mucong Ding, Tahseen Rabbani, Aakriti Agrawal, Yuancheng Xu, Chenghao Deng, Sicheng Zhu, Abdirisak Mohamed, Yuxin Wen, Tom Goldstein, and Furong Huang. WAVES: Benchmarking the Robustness of Image Watermarks. arXiv:2401.08573, 2024. https://arxiv.org/abs/2401.08573.

Mikołaj Bińkowski, Danica J. Sutherland, Michael Arbel, and Arthur Gretton. Demystifying MMD GANs. ICLR, 2018. https://arxiv.org/abs/1801.01401.

Joseph Boutros and Emanuele Viterbo. Signal Space Diversity: A Power- and Bandwidth-Efficient Diversity Technique for the Rayleigh Fading Channel. IEEE Transactions on Information Theory, 44(4):1453–1467, 1998. https://doi.org/10.1109/18.681321.

Hai Ci, Pei Yang, Yiren Song, and Mike Zheng Shou. RingID: Rethinking Tree-Ring Watermarking for Enhanced Multi-Key Identification. ECCV, 2024. https://arxiv.org/abs/2404.14055.

Morris L. Eaton. Group Invariance Applications in Statistics. NSF-CBMS Regional Conference Series in Probability and Statistics, Volume 1. Institute of Mathematical Statistics and American Statistical Association, 1989. https://doi.org/10.1214/cbms/1462061029.

Pierre Fernandez, Guillaume Couairon, Hervé Jégou, Matthijs Douze, and Teddy Furon. The Stable Signature: Rooting Watermarks in Latent Diffusion Models. ICCV, pp. 22466–22477, 2023. https://arxiv.org/abs/2303.15435.

Michel X. Goemans and David P. Williamson. Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming. Journal of the ACM, 42(6):1115–1145, 1995. https://doi.org/10.1145/227683.227684.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. NeurIPS, 2017. https://arxiv.org/abs/1706.08500.

Xiaoxiao Hu, Jiaqi Jin, Sheng Li, Wanli Peng, Xinpeng Zhang, and Zhenxing Qian. Spherical Watermark: Encryption-Free, Lossless Watermarking for Diffusion Models. ICLR, 2026. Conference proceedings.

Aapo Hyvärinen. Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. IEEE Transactions on Neural Networks, 10(3):626–634, 1999. https://www.cs.helsinki.fi/u/ahyvarin/papers/TNN99_reprint.pdf.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver: A Fast ODE Solver for Diffusion Probabilistic Model Sampling in Around 10 Steps. NeurIPS, 2022a. https://arxiv.org/abs/2206.00927.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver++: Fast Solver for Guided Sampling of Diffusion Probabilistic Models. arXiv:2211.01095, 2022b. https://arxiv.org/abs/2211.01095.

Francesco Mezzadri. How to Generate Random Matrices from the Classical Compact Groups. Notices of the American Mathematical Society, 54(5):592–604, 2007. https://arxiv.org/abs/math-ph/0609050.

Art B. Owen. The Pigeonhole Bootstrap. Annals of Applied Statistics, 1(2):386–411, 2007. https://doi.org/10.1214/07-AOAS122.

Henri Poincaré. Science and Hypothesis. Walter Scott Publishing, London, 1905. Chapter 2, Mathematical Magnitude and Experiment. https://brocku.ca/MeadProject/Poincare/Poincare_1905_03.html.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning Transferable Visual Models From Natural Language Supervision. ICML, PMLR 139:8748–8763, 2021. https://proceedings.mlr.press/v139/radford21a.html.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-Resolution Image Synthesis with Latent Diffusion Models. CVPR, pp. 10684–10695, 2022. https://arxiv.org/abs/2112.10752.

Tom Sander, Pierre Fernandez, Alain Durmus, Teddy Furon, and Matthijs Douze. Watermark Anything with Localized Messages. ICLR, 2025. https://arxiv.org/abs/2411.07231.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising Diffusion Implicit Models. ICLR, 2021. https://arxiv.org/abs/2010.02502.

Matthew Tancik, Ben Mildenhall, and Ren Ng. StegaStamp: Invisible Hyperlinks in Physical Photographs. CVPR, pp. 2117–2126, 2020. https://arxiv.org/abs/1904.05343.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. IEEE Transactions on Image Processing, 13(4):600–612, 2004. https://ece.uwaterloo.ca/~z70wang/research/ssim/.

Yuxin Wen, John Kirchenbauer, Jonas Geiping, and Tom Goldstein. Tree-Ring Watermarks: Fingerprints for Diffusion Images that are Invisible and Robust. arXiv:2305.20030, 2023. https://arxiv.org/abs/2305.20030.

Zijin Yang, Kai Zeng, Kejiang Chen, Han Fang, Weiming Zhang, and Nenghai Yu. Gaussian Shading: Provable Performance-Lossless Image Watermarking for Diffusion Models. CVPR, pp. 12162–12171, 2024a. https://arxiv.org/abs/2404.04956.

Pei Yang, Hai Ci, Yiren Song, and Mike Zheng Shou. Can Simple Averaging Defeat Modern Watermarks? NeurIPS, 2024b. Conference proceedings.

Zijin Yang, Xin Zhang, Kejiang Chen, Kai Zeng, Qiyi Yao, Han Fang, Weiming Zhang, and Nenghai Yu. Gaussian Shading++: Rethinking the Realistic Deployment Challenge of Performance-Lossless Image Watermark for Diffusion Models. arXiv:2504.15026, 2025.

Yuexiang Zhai, Zitong Yang, Zhenyu Liao, John Wright, and Yi Ma. Complete Dictionary Learning via ℓ^4-Norm Maximization over the Orthogonal Group. Journal of Machine Learning Research, 21(165):1–68, 2020. https://jmlr.org/papers/v21/19-755.html.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. CVPR, 2018. https://arxiv.org/abs/1801.03924.

Yebin Zheng, Haonan An, Guang Hua, Zhiping Lin, and Yuguang Fang. Correlation-Aware and Gaussianity-Preserving Robust Latent Angular Watermarking for Diffusion Models. arXiv:2607.22386, 2026.

Jiren Zhu, Russell Kaplan, Justin Johnson, and Li Fei-Fei. HiDDeN: Hiding Data with Deep Networks. ECCV, 2018. https://arxiv.org/abs/1807.09937.

SUPPLEMENTARY MATERIAL

Appendices A–C give the implementation, the proofs of the fiber framework and the complete 512-bit evidence, beginning with the channel transplant. The remaining appendices document the earlier studies and broader context.

Appendix at a glance

		page
A Implementation & interfaces	Finite sampling, decoder contract, baseline adaptations and rate accounting	13
B Fiber-framework proofs	Gram invariant, fiber principle, angular law, design on the fiber and Gaussian ladder	17
C Complete 512-bit evidence	Channel transplant, full decoder cells, paired contrasts, diagnostics and resources	24
D Historical 1024-bit study	Separate SD1.5 protocol, numerical matrix and fixed-index gallery	46
E External-state pilot	Earlier dense-frame 512-bit pilot and conditional diagnostic	49
F Reusable-frame analysis	Structural results, failed detector and held-out detector experiments	52
G Historical image evidence	Fixed-index outputs, channel diagnostics and selected failure inspection	58
H Additional related work	Learned and latent watermarks, frame reuse and evaluation	60
I On the epigraph	Poincaré's relations between objects and the fiber view	64

Each earlier study retains its own protocol, dataset and stated rule for selecting image examples.

A IMPLEMENTATION AND BASELINE DETAILS

This appendix gives the modulation and decoder details for the 512-bit comparison. The 1024-bit study uses a separate dataset (Appendix D).

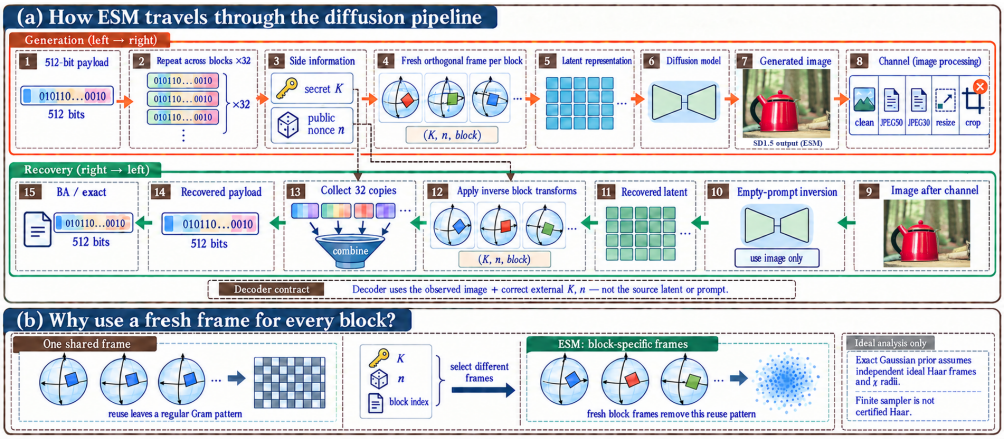


Figure 4: **ESM and what the pipeline sees.** (a) The encoder places repeated payload bits in equal-magnitude codewords and rotates each block in its own frame, regenerated from the key and a per-image nonce. The pipeline acts only on the resulting latent, and the decoder reads the recovered latent through the regenerated frames, combining 32 observations of each bit. Orange and green mark generation and recovery; the thumbnails are an actual SD1.5 output of the study and its JPEG50 version. (b) A shared rotation keeps the inner products between blocks, the Gram invariant of Proposition 1; independent rotations leave only block norms. Exact Gaussianity refers to ideal Haar frames.

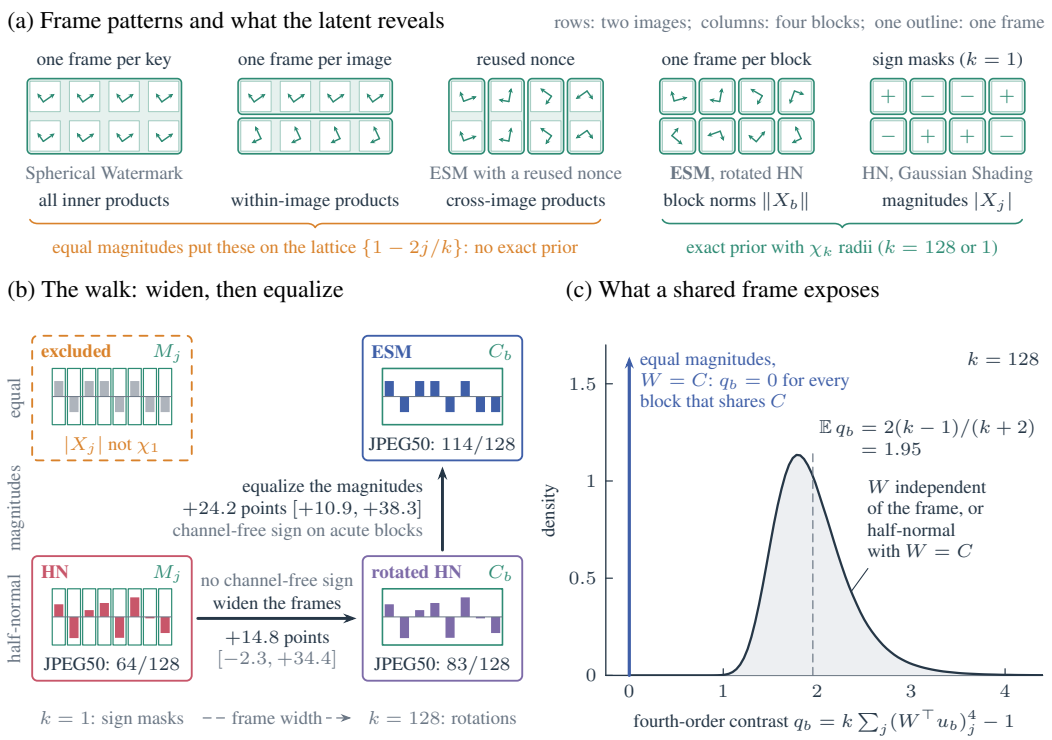


Figure 5: **The decoder asks for equal magnitudes, and equal magnitudes leave one frame pattern.** (a) Frame patterns from Section 3.2 (Table 1; one outline, one frame). Under (A1) the latent reveals only each part's Gram invariant (Proposition 1): sharing exposes inner products, which equal magnitudes put on a lattice. (b) The walk of Table 2 (SD1.5, JPEG50 exact recovery, 95% intervals; bars: a block $S_b \odot \alpha_b$, $k = 8$). Widening the frames has no channel-free sign (Theorem 4(a)); equalizing the magnitudes has one for expected raw sign errors on acute blocks (Theorem 4(c)); the fourth corner violates Proposition 1(i). (c) In its shared frame C, every equal-magnitude block has the minimal contrast $q_b = 0$, against mean $2(k-1)/(k+2)$ for uniform $W^\top u_b$ or half-normal magnitudes (Theorem 8; Monte Carlo, $k = 128$): the flatness that equalizes margins is what a shared frame exposes.

A.1 Factorized ESM

ChaCha20 generates separate pseudorandom streams for each block frame and sign mask from the key and 12-byte image nonce. Frame generation maps an open 52-bit uniform grid through the Gaussian inverse CDF, then normalizes the resulting suffix vectors. Domain separation keeps these draws distinct across blocks and uses.

For a normalized suffix draw $v \in \mathbb{S}^{t-1}$, with $\operatorname{sgn}(0) = +1$, define

$$w = \frac{v + \operatorname{sgn}(v_1) e_1}{\|v + \operatorname{sgn}(v_1) e_1\|}, \qquad H_t = -\operatorname{sgn}(v_1)(I - 2ww^\top).$$

This stable convention satisfies $H_t e_1 = v$. Embedded independent factors give $C = \widetilde{H}_k \cdots \widetilde{H}_1$ by the classical subgroup construction (Mezzadri, 2007). Each factor is symmetric and orthogonal. Encoding applies the factors from the innermost suffix to the outermost, and decoding reverses their order. Independent ideal spherical draws produce a Haar frame; the finite pseudorandom implementation is an approximation to that model.

ESM, HN and rotated-HN share the Gaussian matrix G, 512-bit payload and 32-fold flat repetition within each pair. HN and rotated-HN use $|G|$, while ESM uses $\|G_b\|/\sqrt{128}$ in every coordinate of block b. Their block energies agree, and rotated-HN and ESM also share the frame. The decoder receives the key, nonce, payload size and method parameters; it receives neither the payload nor the original latent or amplitudes.

A.2 Payload layout and decoding

ESM and HN use row-major blocks and a repetition-major payload layout. For zero-based bit $j \in \{0, \dots, p-1\}$, copies occupy flat coordinates $j + p\ell$, $\ell = 0, \dots, r-1$, hence distinct blocks $\lfloor j/k \rfloor + q\ell$. At $p = 512$, $k = 128$ and $r = 32$, bit 0 occupies coordinate 0 in blocks $0, 4, \dots, 124$. This fixed interleaving connects repetition decoding to the ideal law's cross-block independence; the decoder uses no known-message alignment search.

Algorithm 1 Nonce-regenerable ESM encoding and decoding

Encoding inputs: payload m, key K, unique nonce n.
1: Repeat $2m - 1$ r times; reshape into B sign blocks of width k.
2: Draw the Gaussian amplitude vector G and compute each $R_b = \|G_b\|$.
3: Expand the frame-domain stream under (K, n) into ordered suffix draws; normalize each draw and construct its signed Householder factor.
4: Apply factors in reverse suffix-index order to $R_b S_b / \sqrt{k}$; reshape the result into the diffusion latent and generate the image.
Decoding inputs: image, key K, externally supplied nonce n.
5: Invert the image using the fixed empty-prompt pipeline.
6: Recreate the same ordered factors from (K, n); apply them in forward suffix-index order to obtain U.
7: Reshape U into r copies of p coordinates. Apply the fixed majority rule or soft sum, with ties to zero.

Decoder inputs. The image, key and external nonce suffice for decoding; the block frames are regenerated. The angular diagnostic additionally uses the source latent for its conditional calculation.

The surrounding implementation details specify the stream conventions and baseline interfaces. Evaluation compares the decoded bits with the true payload.

A.3 Decoder contract and rate accounting

ESM reconstructs its block frames from the key and externally supplied nonce, then decodes the inverted latent in those frames. The recovery experiments supply the correct image–nonce association. Each baseline receives its prescribed decoder state; Spherical Watermark uses a key-fixed frame.

Nonce uniqueness keeps frame assignments separate across images. Reusing (K, n) recreates the same block frames, so the normalized source blocks from two images satisfy

$$\widetilde{X}_b^\top \widetilde{X}'_b = S_b^\top S'_b / k = 1 - 2d_H(S_b, S'_b)/k.$$

Matching block payloads give collinear directions. This is the cross-image version of the shared-frame relation in Section 3.

The decoder outputs a bit string. Applications that turn this string into a presence, identification or attribution decision also need an acceptance rule and false-positive calibration. The calibration depends on the key/nonce candidates searched.

Table 4: Inputs and outputs of payload decoding, conditional analysis and attribution.

Operation	Required information	Output / scope
Payload decoding	Image, key, correct external nonce	Hard or soft bit string
Angular diagnostic	Original and recovered latents	Conditional ideal-frame prediction
Presence / attribution	Additional decision rule and authenticated association	Not evaluated in this study

The main study carries $512/16384 = 1/32$ useful payload bits per latent coordinate, with a 96-bit ESM nonce carried externally. This rate compares useful payloads at a common latent dimension. Net system rate additionally counts coding, synchronization, authentication and any nonce header. An in-band design would need to recover and locate that header before reconstructing the frames. The experiments use external nonce transport; the earlier 1024-bit study has modulation rate $1/16$.

Regeneration trades retained frame state for repeated computation. The 12-byte nonce, global key and temporary factors are separate resources; seeded dense QR can also regenerate its frames. The historical implementation retained an 8 MiB matrix cache (Appendix D). Reported frame/mask timings exclude nonce transport, lookup, candidate search and authentication.

A.4 Baseline implementations

Gaussian Shading. We adapt the ChaCha20 variant of Gaussian Shading (Yang et al., 2024a) from the authors' implementation[1]. A $2 \times 16 \times 16$ message grid carries 512 bits with $2 \times 4 \times 4 = 32$ spatial/channel repetitions. Encoding preserves the source tensor tiling, bit packing, ChaCha20 XOR and sign assignment; native decoding uses hard majority within the original vote groups, with zero at a 16:16 tie. The paired experiment supplies domain-separated keys and nonces, and float64 absolute-normal samples replace truncated-normal sampling while preserving its conditional distribution. The auxiliary soft decoder removes the sign mask and sums analog coordinates within the same spatial groups. This is a modulation adapter for the original Gaussian Shading method; Gaussian Shading++ uses a different nonce and coding design.

Spherical Watermark. We adapt the modulation and decoding algorithms from the author-supplied archive accompanying Hu et al. (2026). The parameters are $N = 31$, $l_m = l_r = 512$, $t = 1$, $K_{\text{len}} = 128$ and identity padding permutation. Thirty-one message repeats plus 512 fresh padding bits occupy 16384 coordinates, with each message bit using 31 distinct padding positions selected by a permutation. The signs form a 128×128 array; one QR frame multiplies on the left, so blocks are columns and the frame is reused across columns and images for each key, while each column receives an independent χ_{128} radius. Native decoding applies the transpose frame, thresholds signs, estimates padding, undoes the XORs and uses 31-fold majority without receiving the true padding. Study keys generate fresh signets using NumPy PCG64 and QR, preserving the source binary mixing for $t = 1$ and its QR sign convention. The frame is key-fixed, and the common metadata nonce does not change it. The auxiliary soft decoder uses recovered hard padding to flip analog message coordinates before summing the 31 copies. Our common image pipeline and 512-bit payload define the comparison; these measurements are not a reproduction of the complete published system.

LAW and LAW-M. We use paper-based reconstructions of LAW and LAW-M (Zheng et al., 2026), with 512 payload bits and repetition factors $R = 7$ and $R = 1$, respectively. Adjacent Gaussian coordinates form two-dimensional pairs. Each encoding pair keeps its radius and takes an angle of $+\pi/2$ or $-\pi/2$ relative to its reference pair according to the bit. LAW uses the first 3584 pairs for seven contiguous message copies and the next 3584 as references. LAW-M selects the 1024 pairs with largest squared radius before embedding, assigning 512 to encoding and 512 to references; their ordered indices are supplied as per-image decoder side information. For recovered encoding and reference pairs x and r, decoding votes for bit one when $r_1 x_2 - r_2 x_1 < 0$, then takes a majority over the R copies. This signed circular comparison handles angle wrapping at $\pm\pi$. These unkeyed methods reuse the eight request groups of the paired design.

A.5 Models and image pipeline

The common-pipeline comparison uses SD1.5[2] and SD2.1-base[3] with float16 inference at 512×512 resolution. The modulation arrays use float64 and are reshaped into $1 \times 4 \times 64 \times 64$ latents before conversion to float16.

Generation uses 50-step DPM-Solver multistep sampling with guidance 7.5. Inversion uses the VAE posterior mode followed by 50 DDIM steps with an empty prompt and guidance 1. The safety checker remains enabled. JPEG channels use quality 50 or 30. Resize uses Lanczos downsampling to 256×256 and back; crop retains the centered pixel box $[51, 51, 461, 461]$ and resizes it to 512×512.

SW numerical settings. The separate SW study uses 128 paired SD1.5 requests, crossing the same eight test keys and sixteen prompts. Both methods use float32 model weights and inference,

[1] https://github.com/bsmhmmlf/Gaussian-Shading

[2] https://huggingface.co/stable-diffusion-v1-5/stable-diffusion-v1-5

[3] https://huggingface.co/Manojb/stable-diffusion-2-1-base

50-step DPM-Solver generation at guidance 7.5, and sampled VAE posterior latents followed by 50-step empty-prompt DDIM inversion at guidance 1. SW uses the author's float32 Gaussian, QR and chi-square sampling conventions, together with the recovered-padding majority decoder specified above. Fresh signets are derived from the study keys. Posterior-sampling seeds are shared within each pair. We evaluate clean images, JPEG50 and JPEG70, retaining the safety checker.

A.6 Development data and auxiliary analyses

The common TEST grid crosses eight keys with sixteen prompts. For the HN/ESM likelihood check, the disjoint DEV split uses two other keys and four other prompts, giving eight cases per model–method–channel cell. We fit scalar gain and noise level on safe DEV source/recovered pairs and hold them fixed for TEST decoding. TEST decoding uses recovered latents and the fixed parameters, without access to the source image, latent, radius or payload. The scalar-AWGN likelihood and independence across repeats are working assumptions; decoder parameters were fixed before TEST evaluation.

The expanded amplitude ablation compares HN, rotated-HN and ESM on all eight keys and sixteen prompts from SD1.5. Each case shares its payload and Gaussian draw, and rotated-HN and ESM share their frame and execution device. The earlier four-key by eight-prompt comparison is retained separately. These executions are not pooled with the main study. Conditional geometric diagnostics use the original and recovered ESM latents offline and jointly simulate block scores; they are not applied to the shared-frame Spherical Watermark baseline or unrotated HN.

B Proofs for the fiber framework

Throughout, $\mathcal{G} = \prod_{P\in\pi} O(k)$ acts on $(\mathbb{R}^k)^N$ by $(C_P)_P \cdot z = (C_{P(b)} z_b)_b$, and λ denotes Haar probability. Two facts are used repeatedly. First, a compact group acting continuously and transitively on a space admits exactly one invariant probability on it. Second, two tuples $(z_b)_{b\in P}$ and $(z'_b)_{b\in P}$ in $\mathbb{R}^k$ with the same Gram matrix are related by an orthogonal map: the linear map $z_b \mapsto z'_b$ is well defined and isometric on the span, and it extends to $\mathbb{R}^k$.

B.1 Proof of Proposition 1

Proof. Fix a code realization x and let $C \sim \lambda_{\mathcal{G}}$. The law of $C \cdot x$ is $\mathcal{G}$-invariant and carried by the orbit $\mathcal{G}x$, on which $\mathcal{G}$ acts transitively; it is therefore the unique invariant probability $\lambda_{\mathcal{G}x}$. Averaging over the code, $\mathcal{L}(X) = \mathbb{E}_x[\lambda_{\mathcal{G}x}]$. The orbit $\mathcal{G}x$ is the level set $\{z : \Gamma(z) = \Gamma(x)\}$: orthogonal maps preserve inner products, and conversely equal within-part Gram matrices are related part by part by orthogonal maps. Hence $\lambda_{\mathcal{G}x}$ depends on x only through $\Gamma(x)$; write it $\kappa(\Gamma(x), \cdot)$. This is a Markov kernel: $\kappa(\Gamma, \cdot)$ is the law of $C \cdot \tau(\Gamma)$ for any measurable section τ of Γ, which is explicit for parts with $|P| \le k$ (stack $\Gamma_P^{1/2}$ above zero rows) and exists in general by measurable selection. Reflections matter here: with frames drawn from $SO(k)$ instead of $O(k)$, a part with $|P| \ge k$ has the additional invariant $\operatorname{sign}\det$ and the Gram matrix alone no longer determines the orbit. The standard Gaussian is $\mathcal{G}$-invariant, so the same argument gives $N(0, I) = \mathbb{E}_Z[\kappa(\Gamma(Z), \cdot)]$.

Thus $\mathcal{L}(X)$ and $N(0, I)$ are the images of $\mathcal{L}(\Gamma(x))$ and $\mathcal{L}(\Gamma(Z))$ under one kernel, and data processing gives $d_{\mathrm{TV}}(\mathcal{L}(X), N(0, I)) \le d_{\mathrm{TV}}(\mathcal{L}(\Gamma(x)), \mathcal{L}(\Gamma(Z)))$. Conversely, $\Gamma(X) = \Gamma(x)$ because frames preserve inner products, so $\mathcal{L}(\Gamma(x))$ and $\mathcal{L}(\Gamma(Z))$ are images of $\mathcal{L}(X)$ and $N(0, I)$ under the measurable map Γ, which gives the reverse inequality. Distance zero is equality of laws. Under $N(0, I)$ the blocks are independent standard Gaussian vectors, so the $\Gamma_P(Z)$ are independent Gram matrices of $|P|$ such vectors.

For (i), a singleton part has $\Gamma_b = \|x_b\|^2$ and $\|Z_b\| \sim \chi_k$. HN has $|x_j| = |G_j|$, rotated HN has $\|S_b \odot |G_b|\| = \|G_b\|$ and ESM has $\|R_b S_b/\sqrt{k}\| = R_b = \|G_b\|$, whatever the message. For (ii), $X_b^\top X_c = x_b^\top x_c = r_b r_c S_b^\top S_c/k$ and $\|X_b\| = r_b > 0$, so the normalized inner product equals $S_b^\top S_c/k$, which takes $k+1$ values. For independent $Z_b, Z_c \sim N(0, I_k)$ with $k \ge 2$ it has the density proportional to $(1-t^2)^{(k-3)/2}$ on $(-1, 1)$, so the event $\{X_b^\top X_c/(\|X_b\|\|X_c\|) \in \{1 - 2j/k\}_j\}$ has probability one under $\mathcal{L}(X)$ and zero under $N(0, I)$. For (iii), part (ii) excludes, for every message, every pattern with a part of two or more blocks, and for the finest pattern part (i) requires the norms

$\|x_b\| = r_b$ to be independent χ_k variables given each message. The statements are made message by message: after averaging over random payloads the radii need only be χ_k marginally, and for $k = 1$ a mask shared by two coordinates with independent uniform payload bits is exactly Gaussian, which is why (ii) and (iii) assume $k \geq 2$. □

The kernel κ is the classical representation of an invariant law as a mixture of orbit measures indexed by a maximal invariant (Eaton, 1989, Theorems 4.2–4.4). Part (ii) recovers the Gram-lattice identity of Theorem 7; the finite-tolerance bound stated there remains the relevant statement for rounded implementations.

B.2 Proof of Proposition 2

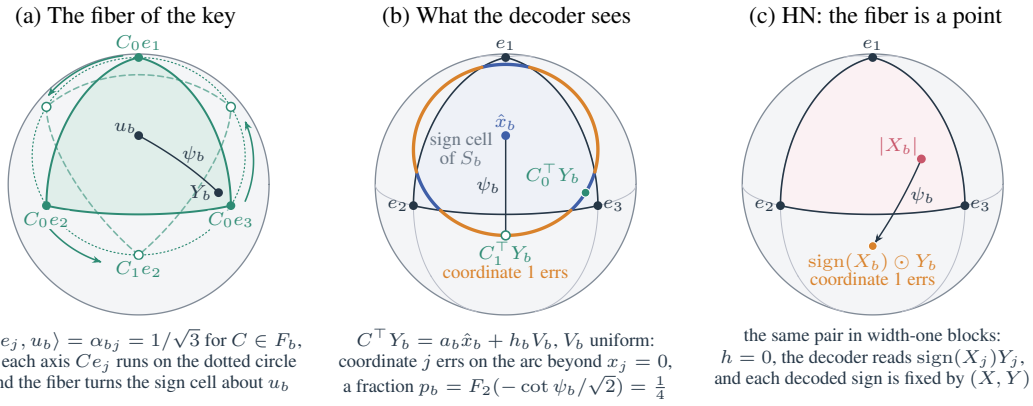


Figure 6: **The transmitted frame is known only up to its fiber, and decoding depends on exactly that.** Proposition 2 under (A1)–(A2), drawn at $k = 3$ with equal magnitudes ($\alpha_{bj} = 1/\sqrt{3}$) and recovered angle $\psi_b = 45°$. (a) Every frame C in the fiber $F_b = \{C \in O(3) : Cx_b = X_b\}$ carries the codeword's axes to the same circle about $u_b = X_b/\|X_b\|$, $\langle Ce_j, u_b\rangle = \alpha_{bj}$ for all $C \in F_b$ (dotted); two members C_0 and $C_1 = R_z(60°)C_0$ turn the same sign cell of S_b to different positions on that circle (arrows), so (x_b, X_b) leaves the frame undetermined. (b) In the codeword's own coordinates, Proposition 2(iii) gives $C^\top Y_b = a_b\hat{x}_b + h_bV_b$ with V_b uniform on the sphere orthogonal to $\hat{x}_b$: as C ranges over F_b, $C^\top Y_b$ traces the circle at angle ψ_b about $\hat{x}_b$, coordinate j errs exactly on the arc beyond $x_j = 0$, and each such arc is a fraction $p_b = F_2(-\cot\psi_b/\sqrt{2}) = \frac{1}{4}$ of the circle (Theorem 3); the two frames of (a) land one inside the sign cell ($C_0^\top Y_b$, correct) and one across its edge ($C_1^\top Y_b$, coordinate 1 errs). (c) For width-one blocks (HN) the fiber is a single point: on the same recovered angle ψ_b, the decoded value $\mathrm{sign}(X_b) \odot Y_b$ is fixed once (X_b, Y_b) are known, with no fiber left to spread it over an arc.

Proof. (i) By Proposition 1, $\mathcal{L}(X) = N(0, I)$ for every message and every prior-preserving code, and by (A2) $Y = \Phi(X, \varepsilon)$ with ε independent of the message, the code and the frames. Hence (X, Y) is obtained from X by one kernel that involves neither the code nor the message.

(ii) Condition on x; the frames are then independent and Haar. Fix b and let T lie in the stabilizer $H_b = \{T \in O(k) : Tx_b = x_b\}$. Right invariance of Haar measure makes C_bT distributed as C_b, and $C_bTx_b = C_bx_b = X_b$. The conditional law of C_b given X_b is therefore invariant under right multiplication by H_b, which acts transitively on the fiber $F_b = C_bH_b$; by uniqueness it is the invariant probability of F_b. Given (x, X), $Y = \Phi(X, \varepsilon)$, and by (A2) ε is independent of (x, C) jointly, hence of the frames given (x, X); conditioning further on Y therefore changes nothing. Parts (ii) and (iii) do not use prior preservation. Independence across blocks follows from the product structure.

(iii) Write $C_b = C_0T$ with $C_0 \in F_b$ fixed and T uniform on H_b. Since $C_b\hat{x}_b = u_b$, $C_b^\top u_b = \hat{x}_b$ and $C_b^\top Y_b = a_b\hat{x}_b + T^\top w$, where $w = C_0^\top(Y_b - a_bu_b)$ is a fixed vector of length h_b in $\hat{x}_b^\perp$. The stabilizer H_b acts on $\hat{x}_b^\perp$ as the full orthogonal group of that $(k-1)$-dimensional space, so $T^\top w$ is uniform on its sphere of radius h_b. □

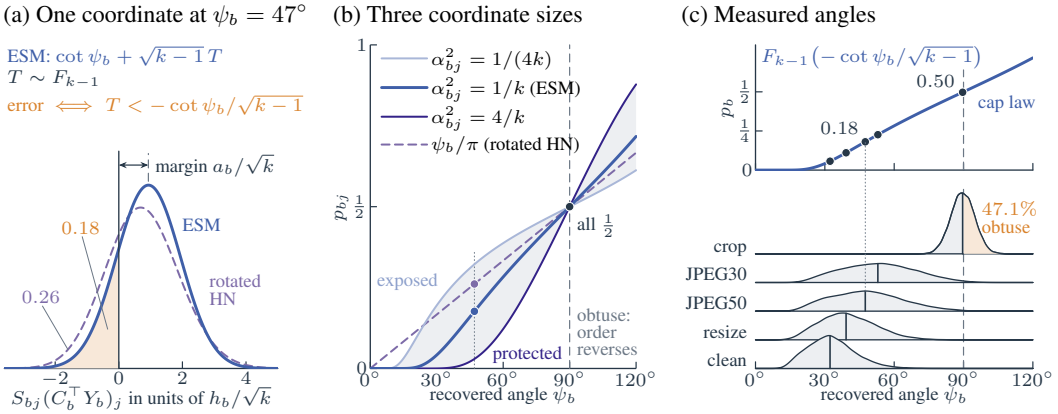


Figure 7: **The recovered angle sets every sign error, and the crop moves it to $90°$.** (a) Theorem 3 under (A1)–(A2), at $k = 128$ and $\psi_b = 47°$: given (x, X, Y), an equal-magnitude coordinate lies a margin $a_b/\sqrt{k}$ from the sign boundary and errs with probability 0.18 (shaded); rotated HN (dashed), the same law averaged over half-normal magnitudes, errs with probability $\psi_b/\pi = 0.26$ (Theorem 4(a)). (b) Equation (2) for $\alpha_{bj}^2 = 1/(4k), 1/k, 4/k$: large coordinates err less on acute blocks and more on obtuse ones, and at $90°$ every coordinate is a fair coin; the half-normal average ψ_b/π lies above the equal-magnitude curve at every acute angle (numerically; Theorem 4(c)); dots repeat (a). (c) Recovered angles of all blocks of the 1280 ESM inversions (SD1.5 and SD2.1-base pooled, 32768 blocks per channel, one density scale), medians marked, beneath the cap law. At the JPEG50 median, $47.5°$, the cap law gives $p_b = 0.18$; after the crop the blocks cluster at the right angle where all curves of (b) meet (median $89.6°$, 47.1% obtuse): a failure of synchronization, not of coding (Section 5.3).

B.3 Proof of Theorem 3

Proof. By Proposition 2(iii), $S_{bj}(C_b^\top Y_b)_j = a_b\alpha_{bj} + h_b\, S_{bj}\langle V_b, e_j\rangle$. Let Π be the orthogonal projection onto $\hat{x}_b^\perp$. Then $\langle V_b, e_j\rangle = \langle V_b, \Pi e_j\rangle$ and $\|\Pi e_j\| = \sqrt{1-\alpha_{bj}^2} > 0$, so $\langle V_b, e_j\rangle = \sqrt{1-\alpha_{bj}^2}\,T_{bj}$, where T_{bj} is the coordinate of a uniform point of the unit sphere of the $(k-1)$-dimensional space $\hat{x}_b^\perp$ along a fixed unit vector; its law is F_{k-1}, which is symmetric, so the sign S_{bj} can be absorbed. The error event is $\{a_b\alpha_{bj} + h_b\sqrt{1-\alpha_{bj}^2}\,T_{bj} < 0\}$, which has the stated probability because $a_b/h_b = \cot\psi_b$; for $k \geq 3$, F_{k-1} has no atoms and ties have probability zero. Distinct blocks use independent conditional frames, so their error events are independent and the error count of a repeated bit has generating function $\prod_b(1 - p_{bj} + p_{bj}t)$. For equal magnitudes, multiplication by $\mathrm{diag}(S_b)$ maps $\hat{x}_b = S_b/\sqrt{k}$ to $\mathbf{1}/\sqrt{k}$ and $\hat{x}_b^\perp$ isometrically onto $\mathbf{1}^\perp$, which gives the stated aligned form. Degenerate cases: if $h_b = 0$ the decoded block is $a_b\hat{x}_b$ and each sign is correct for $a_b > 0$ and wrong for $a_b < 0$; if $Y_b = 0$, tie-to-zero decodes every coordinate as zero. □

B.4 Proof of Theorem 4

Proof. (a) Let M be HN's sign mask and $O = \mathrm{diag}(O_b)$ independent Haar block rotations, both independent of G and of ε by (A1)–(A2). Put $C = O\,\mathrm{diag}(M)$. Right invariance makes C Haar on $O(k)^N$ and independent of the payload, G and ε jointly, because O is independent of all of them; uniformity of M is not needed. HN under the rotated channel transmits $x = \mathrm{diag}(M)(S \odot |G|)$ and decodes

$$\mathrm{diag}(M)O^\top\Phi\big(O\,\mathrm{diag}(M)(S \odot |G|), \varepsilon\big) = C^\top\Phi\big(C(S \odot |G|), \varepsilon\big),$$

which is exactly the analog output of rotated HN with frames C. All decoders are functions of this output and of the payload layout. For the second statement, write $S_b \odot |G_b| = D_bG_b$ with $D_b = \mathrm{diag}(S_b \odot \mathrm{sign}\,G_b)$ and put $C_b' = C_bD_b$. Right invariance makes C_b' Haar and independent of G_b, so $X_b = C_b'G_b$ is independent of C_b', and by (A2) so is (X, Y). With $c = C_b'e_j$, the aligned raw value of coordinate j is $S_{bj}(C_b^\top Y_b)_j = \mathrm{sign}(c^\top X_b)\,c^\top Y_b$, which is negative exactly when the

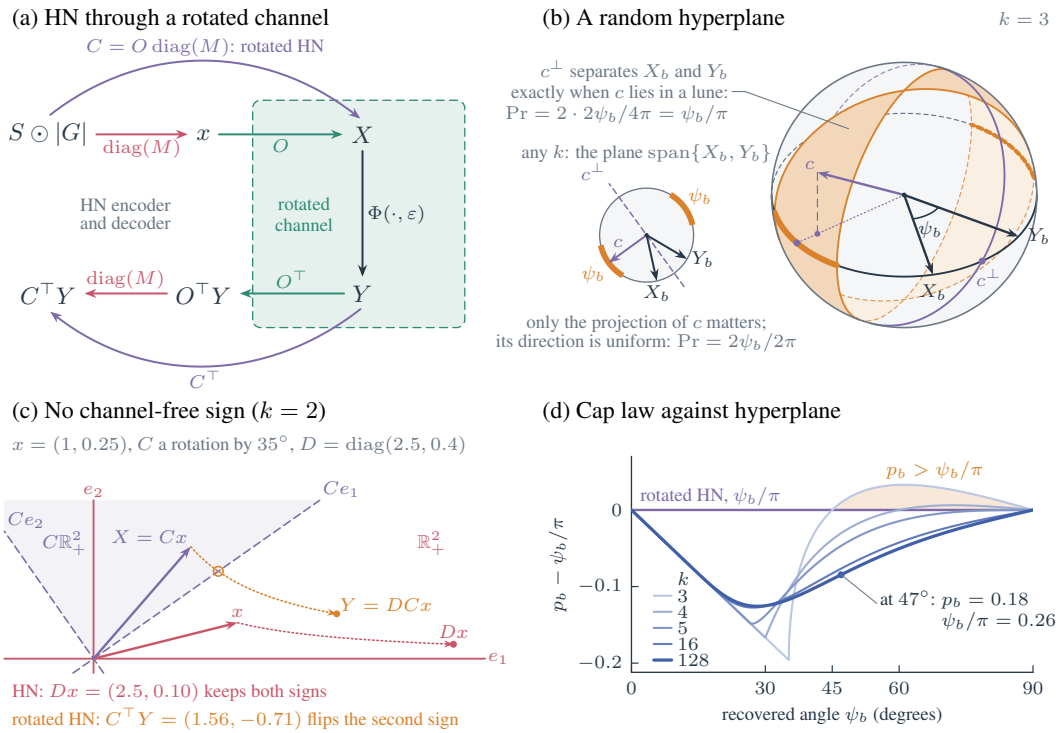


Figure 8: **Independent frames symmetrize the channel, and rotated HN is random-hyperplane rounding.** (a) The identity in the proof of Theorem 4(a), under (A1)–(A2): HN read through the rotated channel $x \mapsto O^\top \Phi(Ox, \varepsilon)$ (box), with independent Haar block rotations O, is rotated HN with frames $C = O\,\mathrm{diag}(M)$, Haar by right invariance. (b) An aligned raw rotated-HN value is $\mathrm{sign}(c^\top X_b)\, c^\top Y_b$ with c uniform and independent of (X, Y); it errs exactly when c lies in one of the two lunes between $X_b^\perp$ and $Y_b^\perp$, with probability ψ_b/π ($\psi_b = 47°$, near the median JPEG50 angle). (c) The positive rescaling D keeps every HN sign but does not commute with the frame: it moves $X = Cx$ out of the rotated sign cell $C\mathbb{R}^2_+$ (dotted: $D^t x$, $D^t X$, $0 \le t \le 1$); on average 15.8% of raw rotated-HN signs err. (d) The cap law p_b of width k minus ψ_b/π: negative at every acute angle for $k = 128$ (numerically), positive on $(45°, 90°)$ for $k = 3$ and on $(60°, 90°)$ for $k = 4$; at $k = 5$ both slopes at $90°$ are $1/\pi$.

hyperplane $c^\perp$ separates X_b from Y_b. For c uniform on the sphere and independent of (X, Y) this has probability ψ_b/π (Goemans & Williamson, 1995).

(b) Proposition 2(iii) gives the aligned block $a_b\alpha_b + h_b\,\mathrm{diag}(S_b)V_b$. The vector V_b has mean zero and $\mathbb{E}\|V_b\|^2 = 1$ when $h_b > 0$, which gives the mean and the trace of the covariance. For $a_b > 0$ the coordinate average of the mean is $a_b k^{-1} \sum_j \alpha_{bj} \le a_b/\sqrt{k}$ by Cauchy–Schwarz with $\sum_j \alpha_{bj}^2 = 1$, with equality exactly when all α_{bj} are equal; for $a_b = 0$ both sides vanish.

(c) By Theorem 3 the conditional expected error count is $\sum_j g(\alpha_{bj}^2)$ with $g(v) = F_{k-1}(-c\sqrt{v/(1-v)})$ and $c = a_b/h_b > 0$. The density of F_{k-1} is $f(t) = c_k(1-t^2)^{(k-4)/2}$ on $(-1, 1)$. Put $x(v) = c\sqrt{v/(1-v)}$, so that $g' = -f(x)x'$ and $g'' = -f'(x)x'^2 - f(x)x''$. With $x' = \frac{c}{2}v^{-1/2}(1-v)^{-3/2}$ and $x'' = \frac{c}{4}v^{-3/2}(1-v)^{-5/2}(4v-1)$, a direct computation on $0 < v < 1/(1+c^2)$, where $x < 1$, gives

$$g''(v) = \frac{c_k\, c}{4}\,(1-x^2)^{(k-6)/2}\, v^{-3/2}(1-v)^{-7/2}\, Q(v), \qquad Q(v) = (1-v)(1-4v) + c^2 v(k-5+4v).$$

For $k \ge 5$ and $v \le 1/4$ both terms of Q are nonnegative and the first is positive for $v < 1/4$. (For $k = 4$ the spread condition admits only equal magnitudes, which is why the statement starts at $k = 5$.) For $v \ge 1/(1+c^2)$, $x \ge 1$ and $g \equiv 0$; at the junction the left derivative is nonpositive and the right derivative is zero. Hence g is convex on $[0, 1/4]$. If $\max_j \alpha_{bj}^2 \le 1/4$, Jensen's

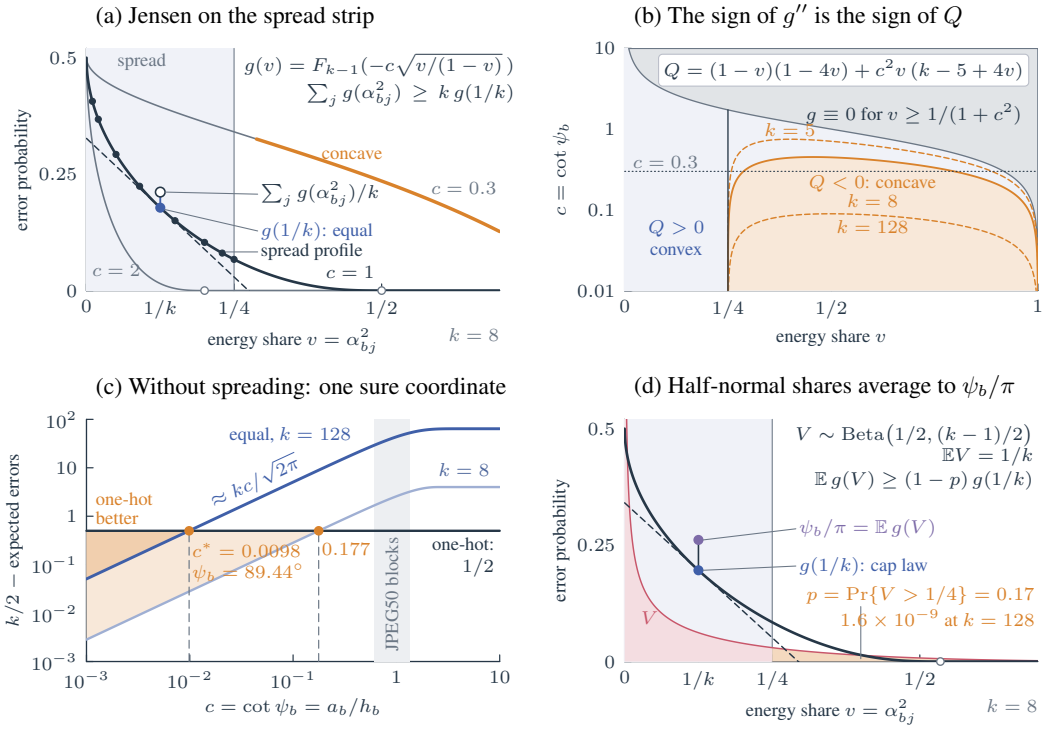


Figure 9: **Equal magnitudes minimize expected sign errors exactly on the spread strip, and bound the cap law by ψ_b/π.** Write $c = \cot\psi_b = a_b/h_b$ and $g(v) = F_{k-1}(-c\sqrt{v/(1-v)})$ for the error probability (2) of a coordinate carrying energy share $v = \alpha_{bj}^2$ (Theorem 3). (a) At $k = 8$, g is convex on the spread strip $v \leq 1/4$ for $c = 0.3, 1, 2$ (Theorem 4(c)), so a spread profile's mean error $\sum_j g(\alpha_{bj}^2)/k = 0.211$ exceeds the equal-magnitude value $g(1/k) = 0.178$ (dots: $\alpha_b^2 = (0.25, 0.23, 0.20, 0.15, 0.09, 0.05, 0.02, 0.01)$); past $v \approx 0.29$ the $c = 0.3$ curve turns concave, which is why the theorem restricts to spread profiles. (b) The sign of g'' is the sign of $Q(v) = (1-v)(1-4v) + c^2 v(k-5+4v)$: both terms are nonnegative on $v \leq 1/4$ for every $c > 0$ once $k \geq 5$, and the concave island $\{Q < 0\}$ (filled, $k = 8$; boundary at $k = 5$ and $k = 128$) lies entirely past that line; $g \equiv 0$ once $v \geq 1/(1+c^2)$. (c) Without spreading, one coordinate can carry the whole margin: the one-hot profile ($k-1$ fair coins, $(k-1)/2$ expected errors) beats the equal-magnitude count $k/2 - kg(1/k)$ once $c < c^*$ (shaded), with $c^* = 0.177$ at $k = 8$ and $c^* = 0.0098$ ($\psi_b = 89.44°$) at $k = 128$ (Theorem 4(c), last claim); the measured interquartile range of c on JPEG50 blocks (ESM, both backbones, 32768 blocks) is $[0.61, 1.34]$, about $60\times$ c^* at $k = 128$. (d) Averaging Theorem 3 over the half-normal profile's own magnitude share $V \sim \mathrm{Beta}(1/2, (k-1)/2)$, $\mathbb{E}V = 1/k$, gives $\mathbb{E}g(V) = \psi_b/\pi$ (Theorem 4(a)); the tangent of g at $1/k$ then bounds the cap law, $g(1/k) \leq (\psi_b/\pi)/(1-p)$ with $p = \Pr\{V > 1/4\}$, tight to a factor $1 + 1.6 \times 10^{-9}$ at $k = 128$ ($k = 8$, $\psi_b = 47°$: $g(1/k) = 0.196$ against $\psi_b/\pi = 0.261$, $p = 0.17$).

inequality with $k^{-1}\sum_j \alpha_{bj}^2 = 1/k \leq 1/4$ gives $\sum_j g(\alpha_{bj}^2) \geq k\, g(1/k)$, the equal-magnitude value. For $a_b < 0$, symmetry of F_{k-1} gives $g_{-c} = 1 - g_c$, which is concave on $[0, 1/4]$, and the inequality reverses. Averaging Theorem 3 over the half-normal profile, whose squared coordinates follow the $\mathrm{Beta}(1/2, (k-1)/2)$ law, and comparing with part (a) gives $\mathbb{E}g(V) = \psi_b/\pi$ for $V \sim \mathrm{Beta}(1/2, (k-1)/2)$, with $\mathbb{E}V = 1/k$. The tangent line ℓ of g at $1/k$ lies below g on $[0, 1/4]$ and is decreasing, so $\mathbb{E}g(V) \geq \mathbb{E}\ell(V) - \mathbb{E}[\ell(V)\mathbf{1}\{V > 1/4\}] \geq g(1/k)\,(1 - \Pr\{V > 1/4\})$: the cap law is at most ψ_b/π divided by $1 - \Pr\{V > 1/4\}$, and $\Pr\{V > 1/4\} = 1.6 \times 10^{-9}$ at $k = 128$. For the last claim, the one-hot profile has one coordinate with $\alpha = 1$, which is never flipped because its decoded value is $a_b > 0$, and $k-1$ coordinates with $\alpha = 0$, each flipped with probability $1/2$; its expected count $(k-1)/2$ is smaller than $k\, g(1/k)$ once $g(1/k) > (k-1)/(2k)$, which holds for all sufficiently small c since $g(1/k) \to 1/2$ as $c \to 0^+$. □

Part (a) has no channel-free sign. A channel that multiplies each latent coordinate by an independent positive factor preserves every sign, so HN makes no error, while rotated HN errs because the factors do not commute with the frames. Numerical checks (supplementary code) confirm the sign of g'' on a grid of $k \in [4, 256]$ and $c \in [10^{-3}, 10^{2}]$, calibrate the general-profile law (2) under a nonlinear, anisotropic channel that mixes blocks, and reproduce part (a) by simulation. They also show that the spread condition is needed: a direct search over unrestricted profiles finds profiles better than equal magnitudes for small a_b/h_b at $k = 4$ and 8. Such blocks are nearly orthogonal to their sources and contribute almost only errors under every profile.

B.5 The Gaussian-noise ladder

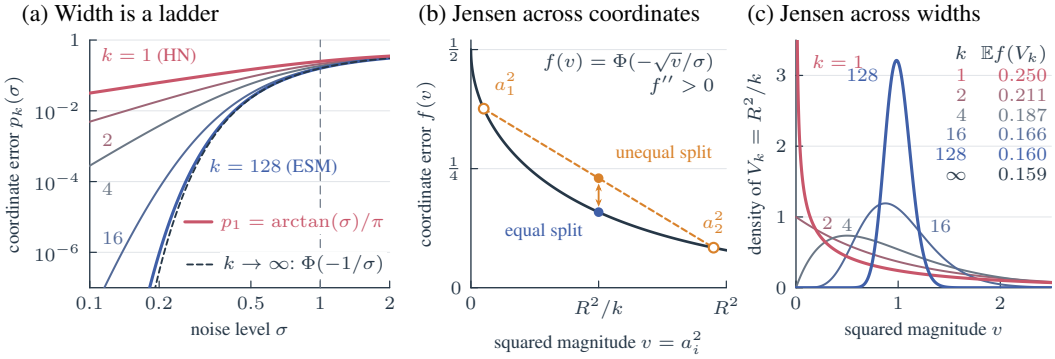


Figure 10: **The Gaussian-noise ladder is Jensen's inequality, used twice.** (a) Under $Y = X + \sigma Z$, the equal-magnitude code of width k with χ_k radii errs on each coordinate with probability $p_k(\sigma) = \Pr\{t_k < -1/\sigma\}$ (Corollary 5). HN is $k = 1$, with the Cauchy tail $\arctan(\sigma)/\pi$, and p_k decreases strictly in k to $\Phi(-1/\sigma)$ (dashed); at $\sigma = 1$ (guide) HN errs with probability $1/4$, ESM ($k = 128$) with 0.160. (b) Given the radius R, magnitudes with $\sum_i a_i^2 = R^2$ err on average $k^{-1}\sum_i f(a_i^2)$ with f convex, so an unequal split (open circles, mean on the chord) errs more than the equal split $a_i^2 = R^2/k$ (blue); $k = 2$, $R^2 = 2$, $\sigma = 1$. (c) At width k every coordinate has squared magnitude $V_k = R^2/k$, a mean of k independent χ_1^2 variables, and $p_k = \mathbb{E}f(V_k)$; a mean of $k+1$ is the average of its $k+1$ leave-one-out means, so strict Jensen gives $\mathbb{E}f(V_{k+1}) < \mathbb{E}f(V_k)$ (table, $\sigma = 1$). The ladder ranks raw coordinate errors; soft repetition sums narrow the HN gap in bit accuracy (Table 3).

Corollary 5 (The Gaussian-noise ladder)**.** *Let $Y = X + \sigma Z$, with $Z \sim N(0, I)$ independent of the encoding. The equal-magnitude code of width k with χ_k radii errs on each coordinate with probability $p_k(\sigma) = \Pr\{t_k < -1/\sigma\}$, the smallest average among all nonnegative magnitudes with the same block energy, where t_k is Student's t variable with k degrees of freedom. HN is the case $k = 1$, with $p_1(\sigma) = \arctan(\sigma)/\pi$, and $p_k(\sigma)$ decreases strictly in k to $\Pr\{\xi < -1/\sigma\}$ for $\xi \sim N(0, 1)$.*

The ladder ranks raw coordinate errors; soft repetition sums narrow the HN gap in bit accuracy (Table 3).

Proof. Additive noise independent of the encoding is frame-blind, and inverse rotation preserves the law of Z and its independence of the encoding variables. Conditional on the radius R, coordinate magnitudes a_i with $\sum_i a_i^2 = R^2$ have average error probability $k^{-1}\sum_i f(a_i^2)$ with $f(v) = \Pr\{\xi < -\sqrt{v}/\sigma\}$, and for $v > 0$

$$f''(v) = \varphi(\sqrt{v}/\sigma)\Big(\frac{1}{4\sigma v^{3/2}} + \frac{1}{4\sigma^3\sqrt{v}}\Big) > 0,$$

where φ is the standard normal density. Continuity extends convexity to zero, and Jensen's inequality gives the unique minimum $a_i = R/\sqrt{k}$ for $R > 0$. Writing $V_k = R^2/k$, the error event is $\xi/\sqrt{V_k} < -1/\sigma$ with $\xi \sim N(0, 1)$ independent of V_k, which is the Student-t tail. For monotonicity, let $Q_1, \ldots, Q_{k+1}$ be independent χ_1^2 variables and write their mean as the average of the $k+1$ leave-one-out means of size k; strict Jensen gives $\mathbb{E}f(V_{k+1}) < \mathbb{E}f(V_k)$. The case $k = 1$ is the Cauchy tail, and the limit follows from the law of large numbers and bounded convergence. For HN, $X_j = M_j S_j |G_j|$ has $|G_j| \sim \chi_1$, which is the case $k = 1$. □

B.6 JOINT CONDITIONAL LAW FOR SOFT AND COMPLETE-MESSAGE DECODING

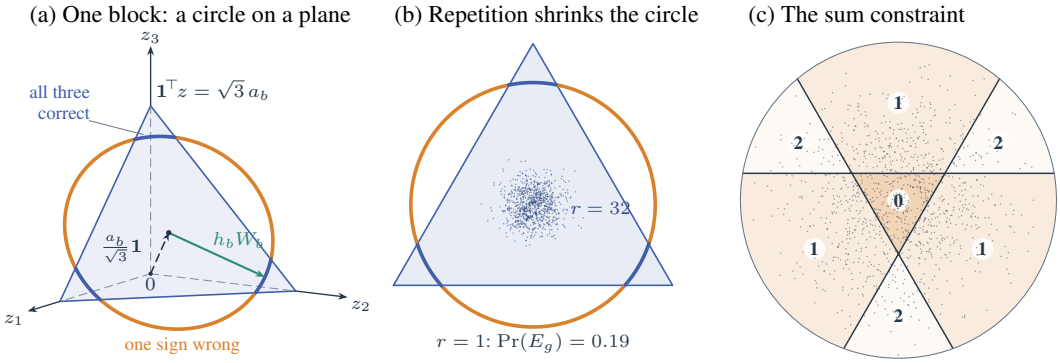


Figure 11: **Group recovery is a point of the plane $\mathbf{1}^\top z = \sqrt{k}\,A_g$ falling in the positive orthant.** Corollary 6 under (A1)–(A2), drawn at $k = 3$. (a) The aligned block $\mathcal{Z}_b$ is uniform on a circle of radius h_b about $(a_b/\sqrt{3})\mathbf{1}$ (3), cut by the orthant in a triangle; at $\psi_b = 47°$ each sign errs with probability $p_b = 0.27$ (Theorem 3), never two at once, so all are correct with probability $1 - 3p_b = 0.19$, not $(1 - p_b)^3 = 0.39$. (b) Scaled by $1/A_g$, the spread of r equal blocks shrinks to $H_g/A_g = \tan\psi_b/\sqrt{r}$; at $k = 128$, A_g/H_g (5.3 here) is each soft sum's mean over deviation, median 4.8 and 5.1 on measured JPEG50 groups. (c) If $A_g \le 0$ the plane misses the open orthant (4): at $A_g/H_g = -0.5$ (crop groups: $[-0.23, 0.30]$) 31% of signs are correct, no cell has three, and the product (6) over groups, not bits, vanishes. After the crop $A_g \le 0$ in 177/512 groups per backbone, so 108/128 and 104/128 messages have success probability zero given (X, Y).

Corollary 6 (Joint signed scores and soft repetition)**.** *Assume $k \ge 3$, equal magnitudes and the conditions of Proposition 2. Retain the definitions of a_b, h_b also when $Y_b = 0$, in which case both are zero. Let $\mathbf{1} \in \mathbb{R}^k$ and $P = I_k - \mathbf{1}\mathbf{1}^\top/k$. Conditional on (X, Y), the signed score vectors satisfy jointly*

$$\mathcal{Z}_b := \operatorname{diag}(S_b)C_b^\top Y_b \overset{d}{=} \frac{a_b}{\sqrt{k}}\mathbf{1} + h_b W_b, \qquad W_b \overset{\text{ind}}{\sim} \operatorname{Unif}(\mathbb{S}^{k-1} \cap \mathbf{1}^\perp). \tag{3}$$

In particular, $\operatorname{Cov}(\mathcal{Z}_b \mid X, Y) = h_b^2 P/(k-1)$: each coordinate has variance h_b^2/k, and distinct coordinates have covariance $-h_b^2/[k(k-1)]$.

For a repetition-major layout with qk payload bits and r repetitions, write $B = qr$ and let group $g = 0, \dots, q-1$ use the distinct blocks $I_g = \{g + q\ell : \ell = 0, \dots, r-1\}$. Their sign vectors are the same payload-sign vector within each group. The vector of its k signed soft sums is $L_g = \sum_{b \in I_g} \mathcal{Z}_b$. Put $A_g = \sum_{b \in I_g} a_b$ and $H_g^2 = \sum_{b \in I_g} h_b^2$. Then

$$L_g \overset{d}{=} \frac{A_g}{\sqrt{k}}\mathbf{1} + \sum_{b \in I_g} h_b W_b, \qquad \mathbf{1}^\top L_g = \sqrt{k} A_g, \qquad \operatorname{Cov}(L_g \mid X, Y) = \frac{H_g^2}{k-1} P. \tag{4}$$

For each coordinate i, its marginal distribution is

$$(L_g)_i \overset{d}{=} \frac{A_g}{\sqrt{k}} + \sqrt{\frac{k-1}{k}} \sum_{b \in I_g} h_b T_b, \qquad T_b \overset{\text{ind}}{\sim} F_{k-1}, \tag{5}$$

with conditional mean $A_g/\sqrt{k}$ and variance H_g^2/k. If $H_g > 0$, the soft bit-error probability is exactly $\Pr\{(L_g)_i < 0 \mid X, Y\}$, computable from this convolution. The variables T_b here describe one coordinate's marginal; drawing them separately for different coordinates would discard the dependence in (3).

Let $m_{g,i}$ be the true payload bit. Under the decoder's tie-to-zero rule, define group success E_g by requiring $(L_g)_i > 0$ when $m_{g,i} = 1$ and $(L_g)_i \ge 0$ when $m_{g,i} = 0$. The disjoint block sets give

$$\Pr\{\widehat{m}^{\text{soft}} = m \mid X, Y\} = \prod_{g=0}^{q-1} \Pr(E_g \mid X, Y). \tag{6}$$

This product is over groups, not over their dependent bit events. If $H_g > 0$, ties have probability zero and $E_g = \{\min_i(L_g)_i > 0\}$ almost surely. If $H_g = 0$, the score is deterministic: $A_g > 0$ makes every bit correct, $A_g < 0$ makes every bit wrong, and $A_g = 0$ makes the decoder output the all-zero group. In particular, a zero recovered block contributes zero to the soft sum, with no universal half-error convention.

Proof. Theorem 3 gives, independently across blocks conditional on (X, Y), $\mathcal{Z}_b = (a_b/\sqrt{k})\mathbf{1} + h_b W_b$ with W_b uniform on the unit sphere of $\mathbf{1}^\perp$ when $h_b > 0$; when $h_b = 0$, an independent W_b may be introduced without affecting the deterministic score. Rotational symmetry in the $(k-1)$-dimensional subspace gives $\mathbb{E}W_b = 0$ and $\mathbb{E}W_b W_b^\top = P/(k-1)$. Within a repetition group the coordinate order and true signs agree across its blocks, so summing these signed vectors is exactly sign alignment of the implemented soft sums. Independence across blocks proves (4). A coordinate's projection onto $\mathbf{1}^\perp$ has norm $\sqrt{(k-1)/k}$, giving (5). For $H_g > 0$, at least one independent summand in each coordinate has a continuous spherical-coordinate law, so its sum is continuous. For $H_g = 0$, every transverse contribution vanishes, yielding the stated deterministic and tie cases. Finally, different L_g depend on disjoint sets of conditionally independent block vectors, so the group events factor as in (6). No within-group bit independence is used. □

Interpretation and scope. The main 512-bit protocol uses $(q, r) = (4, 32)$ at $k = 128$; the historical 1024-bit layout has $(q, r) = (8, 16)$. These indices determine which blocks may be combined independently. For $H_g > 0$ and $A_g \leq 0$, the sum constraint in (4) makes complete group recovery impossible even though individual bits can be correct. Joint simulation can sample each W_b by centering a standard Gaussian vector and normalizing it, thereby preserving all within-block dependence.

This is an offline conditional diagnostic: computing a_b, h_b requires the source latent X, and sign alignment uses the true payload for evaluation. Neither is an additional runtime decoder input. The results of this appendix concern independent ideal frames and channels satisfying (A2); they do not automatically apply to a fixed finite key/nonce sampler or to a reused frame, and they prove neither a Gaussian inversion model nor the optimality of the unweighted sum. Their comparisons between magnitude profiles (Theorem 4) concern expected raw sign errors and the coherent soft sum, not complete-message recovery.

C Additional 512-bit results

C.1 Channel transplant

This re-analysis of stored latents tests Proposition 2(i): if the pipeline's channel is common to all prior-preserving codes, a source/recovered pair measured on one construction's image is a valid channel sample for any other. It uses no generation, inversion or fitting, and its analysis plan was fixed, with its file hashes recorded, before any transplanted outcome was computed; the ordering rests on those hashes and file times, not on an external registry.

Readouts. Every readout acts on a stored source/recovered pair (X, Y) of the main study; rows are blocks, and bit (g, i) uses column i of rows $g+4\ell$, $\ell = 0, \ldots, 31$. The HN readout is deterministic, because the fiber of a width-one frame is a point: the product of mask and payload sign equals $\mathrm{sign}(X)$, so the aligned soft score of bit (g, i) is $\sum_\ell \mathrm{sign}(X_{g+4\ell,i})Y_{g+4\ell,i}$, and a message is recovered when all 512 scores are positive. The ESM readout is the conditional success probability of Corollary 6, computed from the block geometry (a_b, h_b) with 512 joint draws. The rotated-HN readout uses Proposition 2(iii) together with one more fact: under (A1)–(A2), $X_b = C_b x_b$ is independent of the direction $x_b/\|x_b\|$, $\|G_b\|$ is independent of $|G_b|/\|G_b\|$, and Y depends only on X and ε, so the magnitude profile $\alpha_b = |G_b|/\|G_b\|$ is independent of (X, Y) and keeps its prior law. The aligned decoded block is then $a_b\alpha_b + h_b V_b$ with V_b uniform on the unit sphere of $\alpha_b^\perp$; each of 512 joint draws samples fresh profiles and directions, and the soft group scores are summed over the same rows. This conditional law is equivalent to random-hyperplane rounding (Theorem 4(a)). No parameter is fitted. Ties count as errors in the HN and rotated-HN readouts; the ESM readout resolves exact ties as its decoder does, a probability-zero event in the simulation, and only one exact tie occurs among all stored pairs (in a crop cell that already has 256 errors).

Table 5: **One channel, many readouts (JPEG50).** Exact recovery (%) of each readout on channel samples (X, Y) from each construction's images, with no fitted parameter; the last column is the readout's own decoder on its own images. Shaded: own images (HN reproduces its logged outcomes, ESM is the angular-law prediction from a Monte Carlo run independent of Section 5.3). †Separate SD1.5 execution. SW latents are not jointly Gaussian, so that column is descriptive. Intervals for every cell: Tables 6 and 7.

		Channel samples from images of				
Readout	**Model**	**HN**	**GS**	**ESM**	**SW**	**Observed**
HN	SD1.5	53.9	51.6	47.7	50.0	53.9
HN	SD2.1-base	57.8	48.4	50.8	49.2	57.8
Rotated HN	SD1.5	67.5	60.7	63.3	–	64.8†
ESM	SD1.5	89.5	88.6	89.1	86.7	89.1
ESM	SD2.1-base	90.9	87.5	87.4	91.8	85.9

Sources and targets. Channel samples come from the HN, Gaussian Shading and ESM images of the main study, all of which transmit Gaussian latents under ideal masks or frames, on both backbones and all five channels. HN, GS and ESM share the Gaussian draw G of each request: GS uses the same magnitudes $|G|$ as HN under its own sign mask, and ESM has HN's block radii. This correlates predictions with observations within a request and narrows the intervals without biasing the differences. Spherical Watermark pairs are also transplanted and reported descriptively, since its shared frame makes the joint latent non-Gaussian and places it outside Proposition 2. Targets are the observed soft outcomes of HN and ESM in the main study and of rotated HN in the separate SD1.5 mechanism execution (clean and JPEG50), matched by request. The rotated-HN comparison therefore crosses executions: on identical inputs, the two executions disagree on 13 of 128 HN and 6 of 128 ESM JPEG50 outcomes. The agreement statistic is the request-paired difference between the transplanted prediction and the observed outcome, with a 95% crossed-bootstrap interval over keys and prompts (5000 draws). Simulation and bootstrap seeds are SHA-256 digests of the readout, model, case, channel and source labels; the constant recorded in the protocol is not a generator input. The Monte Carlo standard error of a cell mean is about 0.1 to 0.3 points. All cells are reported.

Results. Applied to HN's own 1280 pairs, the HN readout reproduces every logged HN outcome, exact recovery and bit accuracy alike, which checks the implementation. Among the 46 cross-construction cells (Gaussian-latent sources, both backbones, all channels), 44 exact-recovery intervals contain zero. Fourteen of the 46 cells have an observed rate of 0% or 100%, and all 32 cells with interior observed rates contain zero. The two exceptions are ceiling cells: on SD2.1-base under resizing, ESM recovers 128/128 messages, while the transplanted predictions are 99.2% from HN images and 99.8% from GS images, differences of -0.81 and -0.23 points whose intervals end just below zero; under the predicted probabilities, 128/128 successes have probability about 0.35 and 0.75, and the GS-source cell falls inside its interval for most Monte Carlo seeds. Over the JPEG and resize channels the mean absolute difference is 2.8 points; the largest, 10.2 points (SD2.1-base, JPEG30), lies inside its interval. Bit-accuracy differences are at most 0.49 points in absolute value, and the same two ceiling cells are the only cross-construction intervals that exclude zero. The twenty cells with Spherical Watermark sources contain zero in exact recovery, and 19 of 20 in bit accuracy (the exception is SD1.5 resize for the ESM readout, -0.008 points), although that construction lies outside the premise.

Two patterns deserve mention. All transplanted HN rates under JPEG compression lie below HN's observed rate, by 2.3 to 10.2 points. These differences share one observed reference per model and channel, so they are not independent; on SD1.5 the separate mechanism execution measured 50.0% for HN under JPEG50, inside the transplanted range of 47.7–51.6%, against 53.9% in the main execution. On SD2.1-base the same sign appears in bit accuracy (-0.04 to -0.33 points, all inside their intervals), and no second SD2.1-base execution exists to test the shared-reference explanation. Second, reading down any column of Table 5 compares readouts of identical channel samples, which removes image-to-image variation: the order HN $<$ rotated HN $<$ ESM holds in every column where all three readouts are available, for instance 47.7%, 63.3% and 89.1% on the ESM images of SD1.5. These within-column comparisons are descriptive and were not part of the fixed protocol.

Table 8 reports the recovered block angles. Under JPEG50, 99.3–99.8% of blocks are acute, as Theorem 4(c) requires; after the crop the median angle is within half a degree of a right angle and 47–48% of blocks are obtuse.

Table 6: **All transplant cells, SD1.5.** Predicted / observed exact recovery (%) and the request-paired difference in points with its 95% crossed-bootstrap interval. Rows marked 'own' use the readout's own images.

Channel	Readout	Samples from	Pred. / obs.	Δ [95% CI]
Clean	HN	HN (own)	96.88 / 96.88	+0.00 [+0.00, +0.00]
Clean	HN	GS	96.88 / 96.88	+0.00 [−5.47, +6.25]
Clean	HN	ESM	94.53 / 96.88	−2.34 [−8.59, +2.34]
Clean	HN	SW	94.53 / 96.88	−2.34 [−10.94, +3.12]
Clean	Rotated HN	HN	98.45 / 97.66	+0.79 [−2.76, +6.30]
Clean	Rotated HN	GS	98.48 / 97.66	+0.83 [−1.88, +6.18]
Clean	Rotated HN	ESM	97.51 / 97.66	−0.15 [−4.70, +4.13]
Clean	**ESM**	HN	99.94 / 100.00	−0.06 [−0.27, +0.00]
Clean	**ESM**	GS	99.96 / 100.00	−0.04 [−0.26, +0.00]
Clean	**ESM**	ESM (own)	99.87 / 100.00	−0.13 [−0.55, +0.00]
Clean	**ESM**	SW	99.89 / 100.00	−0.11 [−0.42, +0.00]
JPEG50	HN	HN (own)	53.91 / 53.91	+0.00 [+0.00, +0.00]
JPEG50	HN	GS	51.56 / 53.91	−2.34 [−18.75, +13.28]
JPEG50	HN	ESM	47.66 / 53.91	−6.25 [−22.66, +9.38]
JPEG50	HN	SW	50.00 / 53.91	−3.91 [−22.66, +14.84]
JPEG50	Rotated HN	HN	67.46 / 64.84	+2.62 [−12.26, +15.86]
JPEG50	Rotated HN	GS	60.67 / 64.84	−4.17 [−17.66, +7.86]
JPEG50	Rotated HN	ESM	63.26 / 64.84	−1.58 [−15.56, +12.15]
JPEG50	**ESM**	HN	89.52 / 89.06	+0.46 [−9.01, +11.11]
JPEG50	**ESM**	GS	88.56 / 89.06	−0.51 [−9.74, +9.77]
JPEG50	**ESM**	ESM (own)	89.12 / 89.06	+0.06 [−4.77, +5.46]
JPEG50	**ESM**	SW	86.69 / 89.06	−2.38 [−9.67, +4.99]
JPEG30	HN	HN (own)	38.28 / 38.28	+0.00 [+0.00, +0.00]
JPEG30	HN	GS	35.94 / 38.28	−2.34 [−17.97, +12.50]
JPEG30	HN	ESM	34.38 / 38.28	−3.91 [−19.53, +10.94]
JPEG30	HN	SW	30.47 / 38.28	−7.81 [−21.09, +4.69]
JPEG30	**ESM**	HN	75.66 / 75.78	−0.12 [−14.61, +14.52]
JPEG30	**ESM**	GS	71.77 / 75.78	−4.01 [−15.08, +7.43]
JPEG30	**ESM**	ESM (own)	72.56 / 75.78	−3.22 [−9.67, +3.12]
JPEG30	**ESM**	SW	72.45 / 75.78	−3.33 [−13.95, +7.36]
Resize	HN	HN (own)	82.81 / 82.81	+0.00 [+0.00, +0.00]
Resize	HN	GS	83.59 / 82.81	+0.78 [−13.28, +16.41]
Resize	HN	ESM	83.59 / 82.81	+0.78 [−14.84, +16.41]
Resize	HN	SW	83.59 / 82.81	+0.78 [−14.06, +16.41]
Resize	**ESM**	HN	97.98 / 97.66	+0.32 [−3.13, +5.28]
Resize	**ESM**	GS	98.22 / 97.66	+0.56 [−1.75, +4.39]
Resize	**ESM**	ESM (own)	97.24 / 97.66	−0.42 [−3.60, +1.92]
Resize	**ESM**	SW	96.09 / 97.66	−1.57 [−5.85, +0.63]
Crop	HN	HN (own)	0.00 / 0.00	+0.00 [+0.00, +0.00]
Crop	HN	GS	0.00 / 0.00	+0.00 [+0.00, +0.00]
Crop	HN	ESM	0.00 / 0.00	+0.00 [+0.00, +0.00]
Crop	HN	SW	0.00 / 0.00	+0.00 [+0.00, +0.00]
Crop	**ESM**	HN	0.00 / 0.00	+0.00 [+0.00, +0.00]
Crop	**ESM**	GS	0.00 / 0.00	+0.00 [+0.00, +0.00]
Crop	**ESM**	ESM (own)	0.00 / 0.00	+0.00 [+0.00, +0.00]
Crop	**ESM**	SW	0.00 / 0.00	+0.00 [+0.00, +0.00]

Table 7: **All transplant cells, SD2.1-base.** Predicted / observed exact recovery (%) and the request-paired difference in points with its 95% crossed-bootstrap interval. Rows marked 'own' use the readout's own images.

Channel	Readout	Samples from	Pred. / obs.	Δ [95% CI]
Clean	HN	HN (own)	97.66 / 97.66	$+0.00$ $[+0.00, +0.00]$
Clean	HN	GS	98.44 / 97.66	$+0.78$ $[-5.47, +6.25]$
Clean	HN	ESM	97.66 / 97.66	$+0.00$ $[-6.25, +4.69]$
Clean	HN	SW	96.09 / 97.66	-1.56 $[-9.38, +4.69]$
Clean	**ESM**	HN	99.99 / 100.00	-0.01 $[-0.05, +0.00]$
Clean	**ESM**	GS	100.00 / 100.00	$+0.00$ $[+0.00, +0.00]$
Clean	**ESM**	ESM (own)	100.00 / 100.00	$+0.00$ $[+0.00, +0.00]$
Clean	**ESM**	SW	100.00 / 100.00	$+0.00$ $[+0.00, +0.00]$
JPEG50	HN	HN (own)	57.81 / 57.81	$+0.00$ $[+0.00, +0.00]$
JPEG50	HN	GS	48.44 / 57.81	-9.38 $[-28.12, +9.38]$
JPEG50	HN	ESM	50.78 / 57.81	-7.03 $[-28.12, +11.72]$
JPEG50	HN	SW	49.22 / 57.81	-8.59 $[-24.22, +6.25]$
JPEG50	**ESM**	HN	90.90 / 85.94	$+4.96$ $[-4.96, +16.12]$
JPEG50	**ESM**	GS	87.47 / 85.94	$+1.53$ $[-6.49, +10.07]$
JPEG50	**ESM**	ESM (own)	87.36 / 85.94	$+1.42$ $[-3.73, +8.46]$
JPEG50	**ESM**	SW	91.78 / 85.94	$+5.84$ $[-4.19, +18.70]$
JPEG30	HN	HN (own)	39.06 / 39.06	$+0.00$ $[+0.00, +0.00]$
JPEG30	HN	GS	28.91 / 39.06	-10.16 $[-23.44, +2.34]$
JPEG30	HN	ESM	36.72 / 39.06	-2.34 $[-16.41, +10.94]$
JPEG30	HN	SW	32.03 / 39.06	-7.03 $[-21.88, +7.03]$
JPEG30	**ESM**	HN	76.32 / 76.56	-0.24 $[-11.73, +10.45]$
JPEG30	**ESM**	GS	73.39 / 76.56	-3.17 $[-13.09, +7.33]$
JPEG30	**ESM**	ESM (own)	77.42 / 76.56	$+0.86$ $[-4.24, +7.16]$
JPEG30	**ESM**	SW	76.42 / 76.56	-0.14 $[-14.23, +15.44]$
Resize	HN	HN (own)	85.16 / 85.16	$+0.00$ $[+0.00, +0.00]$
Resize	HN	GS	86.72 / 85.16	$+1.56$ $[-10.94, +14.06]$
Resize	HN	ESM	89.84 / 85.16	$+4.69$ $[-7.03, +16.41]$
Resize	HN	SW	85.16 / 85.16	$+0.00$ $[-12.50, +12.50]$
Resize	**ESM**	HN	99.19 / 100.00	-0.81 $[-2.27, -0.04]$
Resize	**ESM**	GS	99.77 / 100.00	-0.23 $[-0.80, -0.003]$
Resize	**ESM**	ESM (own)	99.78 / 100.00	-0.22 $[-0.82, +0.00]$
Resize	**ESM**	SW	99.74 / 100.00	-0.26 $[-0.81, +0.00]$
Crop	HN	HN (own)	0.00 / 0.00	$+0.00$ $[+0.00, +0.00]$
Crop	HN	GS	0.00 / 0.00	$+0.00$ $[+0.00, +0.00]$
Crop	HN	ESM	0.00 / 0.00	$+0.00$ $[+0.00, +0.00]$
Crop	HN	SW	0.00 / 0.00	$+0.00$ $[+0.00, +0.00]$
Crop	**ESM**	HN	0.00 / 0.00	$+0.00$ $[+0.00, +0.00]$
Crop	**ESM**	GS	0.00 / 0.00	$+0.00$ $[+0.00, +0.00]$
Crop	**ESM**	ESM (own)	0.00 / 0.00	$+0.00$ $[+0.00, +0.00]$
Crop	**ESM**	SW	0.00 / 0.00	$+0.00$ $[+0.00, +0.00]$

C.2 Outer-code re-analysis

All methods in the main study protect the payload by 32-fold repetition only, so their exact-recovery rates could overstate the differences that remain after a standard outer code. This post hoc CPU re-analysis, motivated by that concern, decodes every stored recovered latent with the method's own native decoder, forms the 512-bit error pattern, and evaluates bounded-distance Reed–Solomon decoding over GF(2^8) on the 64 consecutive bytes, correcting t symbol errors. Success depends on the error pattern only, so the evaluation is exact for such a decoder; it idealizes the embedding,

Table 8: **Recovered block angles.** Angle ψ_b between transmitted and recovered blocks over all 128×128 blocks per cell, measured on ESM and HN images. Obtuse blocks ($\psi_b > 90°$) reverse the equal-margin inequality of Theorem 4(c).

Model	Channel	Images	Median	99th pct.	Max	Obtuse (%)
SD1.5	Clean	ESM	32.6°	61.5°	79.3°	0.00
SD1.5	Clean	HN	32.4°	59.6°	79.8°	0.00
SD1.5	JPEG50	ESM	48.2°	84.6°	97.9°	0.20
SD1.5	JPEG50	HN	46.7°	84.3°	101.3°	0.24
SD1.5	JPEG30	ESM	53.8°	88.4°	102.1°	0.67
SD1.5	JPEG30	HN	52.1°	88.2°	102.9°	0.60
SD1.5	Resize	ESM	39.8°	72.0°	97.5°	0.02
SD1.5	Resize	HN	39.3°	70.6°	88.0°	0.00
SD1.5	Crop	ESM	89.6°	101.4°	109.2°	47.24
SD1.5	Crop	HN	89.7°	101.8°	108.6°	47.30
SD2.1-base	Clean	ESM	31.8°	57.7°	80.1°	0.00
SD2.1-base	Clean	HN	32.0°	62.5°	104.1°	0.12
SD2.1-base	JPEG50	ESM	46.7°	86.7°	100.7°	0.45
SD2.1-base	JPEG50	HN	46.6°	88.2°	102.5°	0.68
SD2.1-base	JPEG30	ESM	52.1°	88.9°	103.3°	0.76
SD2.1-base	JPEG30	HN	52.3°	89.5°	102.8°	0.92
SD2.1-base	Resize	ESM	38.6°	68.1°	86.4°	0.00
SD2.1-base	Resize	HN	38.8°	72.8°	100.2°	0.15
SD2.1-base	Crop	ESM	89.6°	101.5°	111.6°	47.04
SD2.1-base	Crop	HN	89.7°	101.8°	107.3°	48.13

because the embedded bits were random payloads rather than codewords, which is equivalent under the symmetric ideal model. The check $t = 0$ reproduces the logged exact outcome of all 5120 records.

Table 9 reports every cell. The outer code narrows the gaps but does not close them. Under JPEG50 at $t = 2$ (480 net bits), ESM recovers 96.9% and 93.0% of messages on SD1.5 and SD2.1-base against 69.5% and 73.4% for HN; at $t = 8$ (384 net bits), 99.2% and 97.7% against 86.7% and 93.0%, where the SD2.1-base interval includes zero. Under JPEG30 the ESM–HN gap at $t = 8$ is 24.2 and 14.8 points, and both intervals exclude zero. The crop leaves every method at zero for every t.

C.3 Interleaving versus rotation

HN's repetition-major layout places the 32 copies of bit j at flat coordinates $j + 512\ell$, all in latent column $j \bmod 64$, whereas a frame of width 128 spreads each block over two full latent rows. The step from HN to rotated HN could therefore mix orientation averaging with plain spatial interleaving. This post hoc re-analysis, motivated by that concern, separates the two on identical channel samples. HN with a uniformly random permutation of the 16384 coordinates before its layout is still a prior-preserving width-one code, and its readout is deterministic given (X, Y); we average its success over 256 random interleavers per stored pair. A random interleaver is itself a symmetrization, by the permutation group rather than by rotations.

Table 10 reports every cell. Interleaving accounts for most of the HN-to-rotated-HN step: under JPEG50 on SD1.5 it raises HN's exact recovery by 7.4 to 13.7 points on the three sources, and rotation adds 1.5 to 1.9 points beyond it, with intervals that exclude zero. The gains are similar on SD2.1-base (7.9 to 17.4 points) and under JPEG30. Against interleaved HN, ESM still recovers 23.6 to 29.6 more JPEG50 messages per hundred on SD1.5 and 19.2 to 25.2 on SD2.1-base, every interval excluding zero, so the equal-magnitude step, not the layout, carries the headline difference.

C.4 Evaluation design

The common-pipeline studies evaluate 1536 generated images under five channels: 1024 from the four-method main comparison and 512 from the exploratory LAW/LAW-M reconstructions. Each

Table 9: **Complete-message recovery with an idealized outer code.** Success rate (%) of bounded-distance Reed–Solomon decoding over GF(2^8) that corrects t of the 64 byte symbols, evaluated on the measured error pattern of each request's native decoder; net payload bits in parentheses ($t = 1$ is in the released results). $t = 0$ reproduces the logged exact recovery of all 5120 records. The last two columns give ESM minus the row's method at $t = 2$ and $t = 8$, in points, with 95% crossed-bootstrap intervals.

Channel	Method	$t = 0$ (512)	$t = 2$ (480)	$t = 4$ (448)	$t = 8$ (384)	$\Delta, t = 2$	$\Delta, t = 8$
SD1.5							
JPEG50	HN	53.9	69.5	78.1	86.7	+27.3 [+11.7, +44.5]	+12.5 [+3.1, +25.8]
	GS	35.9	51.6	64.1	78.1	+45.3 [+28.1, +62.5]	+21.1 [+7.8, +35.9]
	SW	27.3	32.8	36.7	45.3	+64.1 [+44.5, +80.5]	+53.9 [+35.2, +71.9]
	ESM	89.1	96.9	99.2	99.2		
JPEG30	HN	38.3	50.8	60.9	72.7	+34.4 [+18.0, +51.6]	+24.2 [+10.2, +40.6]
	GS	26.6	39.1	47.7	60.2	+46.1 [+30.5, +62.5]	+36.7 [+20.3, +54.7]
	SW	14.8	21.9	28.1	31.2	+63.3 [+46.1, +78.9]	+65.6 [+46.1, +83.6]
	ESM	75.8	85.2	93.0	96.9		
Resize	HN	82.8	95.3	98.4	98.4	+3.1 [+0.0, +10.9]	+1.6 [+0.0, +7.0]
	GS	66.4	86.7	93.0	96.1	+11.7 [+2.3, +25.0]	+3.9 [+0.0, +14.1]
	SW	50.0	71.9	76.6	83.6	+26.6 [+11.7, +44.5]	+16.4 [+4.7, +31.2]
	ESM	97.7	98.4	100.0	100.0		
SD2.1-base							
JPEG50	HN	57.8	73.4	84.4	93.0	+19.5 [+5.5, +35.2]	+4.7 [−3.9, +14.1]
	GS	35.2	58.6	68.8	79.7	+34.4 [+20.3, +49.2]	+18.0 [+6.2, +32.8]
	SW	27.3	37.5	39.1	44.5	+55.5 [+36.7, +74.2]	+53.1 [+33.6, +72.7]
	ESM	85.9	93.0	93.8	97.7		
JPEG30	HN	39.1	53.9	63.3	78.9	+28.9 [+13.3, +45.3]	+14.8 [+3.9, +28.1]
	GS	25.8	37.5	48.4	61.7	+45.3 [+27.3, +63.3]	+32.0 [+18.0, +47.7]
	SW	17.2	21.9	24.2	29.7	+60.9 [+42.2, +78.9]	+64.1 [+43.0, +82.8]
	ESM	76.6	82.8	85.9	93.8		
Resize	HN	85.2	94.5	98.4	98.4	+5.5 [+0.0, +14.1]	+1.6 [+0.0, +6.2]
	GS	66.4	89.8	96.9	98.4	+10.2 [+0.8, +22.7]	+1.6 [+0.0, +7.0]
	SW	53.1	71.9	79.7	85.9	+28.1 [+13.3, +45.3]	+14.1 [+4.7, +25.8]
	ESM	100.0	100.0	100.0	100.0		

model–method pair has eight request groups and sixteen prompts. The expanded amplitude ablation uses 128 fresh three-method paired executions; the historical 32-pair ablation remains separate. These studies have no safety-filtered images. The earlier 1024-bit study uses its own dataset and appears in Appendix D.

C.5 Exploratory LAW and LAW-M reconstructions

The LAW and LAW-M reconstructions use the same 128 request descriptors per backbone as the original common-pipeline study, with the implementation choices specified in Appendix A.4. We report them separately because their fidelity to the authors' system remains unverified. The source evaluates bit accuracy and thresholded detection on COCO captions; here we also count complete 512-bit messages on authored prompts. LAW uses seven repetitions and LAW-M one, with 1024 ordered pair indices supplied for LAW-M decoding. Their encoding and reference pairs share the latent budget, so repetition counts alone do not describe the rate comparison. All 512 generation requests pass the safety checker. Tables 11 and 12 report every method–channel cell, including pointwise 95% intervals from crossed resampling of request groups and prompts.

The secondary comparisons match each new baseline result to the original ESM request by model, group, prompt, payload and generation descriptor. They compare separate executions under the common pipeline. Tables 13 and 14 report all paired BA and exact-recovery differences, keeping 128 requests per arm and using 5000 crossed-bootstrap draws.

Under JPEG50, ESM's BA advantage over LAW and LAW-M is 16.01 and 13.70 points on SD1.5, and 15.17 and 15.08 points on SD2.1-base. Each difference is positive in all eight request groups. LAW recovers no complete JPEG50 message on either model; LAW-M recovers none on SD1.5 and two on SD2.1-base, compared with ESM's 114 and 110. Under the centered crop, every method has zero complete recoveries. These measurements evaluate the declared paper reconstructions, including LAW-M's supplied pair indices.

Table 10: **Interleaving versus rotation on identical channel samples.** Exact recovery (%) of four readouts of the same stored pairs: HN with its layout, HN with a uniformly random interleaver (average over 256 interleavers), rotated HN (SD1.5 JPEG50 only) and ESM. The last two columns give interleaved minus layout HN and ESM minus interleaved HN, in points, with 95% crossed-bootstrap intervals. Post hoc analysis.

Channel	Samples	HN	Interleaved HN	Rotated HN	ESM	Interleaved − HN	ESM − interleaved
SD1.5							
JPEG50	HN images	53.9	66.0	67.5	89.5	$+12.1\,[+1.5, +24.0]$	$+23.6\,[+12.8, +35.0]$
	GS images	51.6	59.0	60.7	88.6	$+7.4\,[-2.8, +18.6]$	$+29.6\,[+19.1, +40.1]$
	ESM images	47.7	61.4	63.3	89.1	$+13.7\,[+5.1, +23.7]$	$+27.7\,[+15.9, +39.9]$
JPEG30	HN images	38.3	49.3	–	75.7	$+11.1\,[+0.5, +21.9]$	$+26.3\,[+16.5, +37.0]$
	GS images	35.9	42.8	–	71.8	$+6.8\,[-0.7, +16.0]$	$+29.0\,[+18.9, +39.6]$
	ESM images	34.4	44.1	–	72.6	$+9.7\,[+1.7, +19.9]$	$+28.5\,[+18.7, +39.0]$
Resize	HN images	82.8	91.7	–	98.0	$+8.9\,[+0.9, +18.7]$	$+6.3\,[+1.9, +14.2]$
	GS images	83.6	90.7	–	98.2	$+7.1\,[+1.2, +14.1]$	$+7.5\,[+2.0, +16.2]$
	ESM images	83.6	90.8	–	97.2	$+7.2\,[+0.2, +16.6]$	$+6.5\,[+2.1, +13.2]$
SD2.1-base							
JPEG50	HN images	57.8	65.7	–	90.9	$+7.9\,[-1.7, +19.0]$	$+25.2\,[+13.5, +37.6]$
	GS images	48.4	63.0	–	87.5	$+14.6\,[+4.5, +26.1]$	$+24.4\,[+16.1, +33.5]$
	ESM images	50.8	68.2	–	87.4	$+17.4\,[+4.3, +31.6]$	$+19.2\,[+11.2, +27.9]$
JPEG30	HN images	39.1	48.5	–	76.3	$+9.5\,[+1.1, +19.6]$	$+27.8\,[+17.2, +38.5]$
	GS images	28.9	42.8	–	73.4	$+13.8\,[+4.7, +24.4]$	$+30.6\,[+20.6, +41.4]$
	ESM images	36.7	50.6	–	77.4	$+13.9\,[+5.2, +24.2]$	$+26.8\,[+16.3, +39.3]$
Resize	HN images	85.2	91.6	–	99.2	$+6.4\,[-0.5, +15.6]$	$+7.6\,[+2.3, +14.5]$
	GS images	86.7	94.0	–	99.8	$+7.3\,[+0.8, +15.0]$	$+5.8\,[+1.7, +12.5]$
	ESM images	89.8	95.1	–	99.8	$+5.3\,[-0.4, +13.1]$	$+4.7\,[+1.0, +11.0]$

Table 11: Paper-based LAW reconstruction recovery on SD1.5. Delivered BA and interval bounds are percentages; exact recovery is shown as count/128. Brackets give pointwise 95% crossed-bootstrap intervals.

Method	Channel	BA (%) [95% CI]	Exact/128 [rate CI (%)]
SD1.5			
LAW ($R = 7$)	Clean	$96.30\ [93.98, 97.95]$	$8/128\ [0.00, 17.19]$
LAW ($R = 7$)	JPEG50	$83.81\ [78.97, 88.58]$	$0/128\ [0.00, 0.00]$
LAW ($R = 7$)	JPEG30	$78.57\ [73.08, 84.21]$	$0/128\ [0.00, 0.00]$
LAW ($R = 7$)	Resize256	$91.55\ [88.03, 94.57]$	$1/128\ [0.00, 4.69]$
LAW ($R = 7$)	Crop410	$50.07\ [49.43, 50.71]$	$0/128\ [0.00, 0.00]$
LAW-M ($R = 1$)	Clean	$97.20\ [95.21, 98.52]$	$10/128\ [0.78, 18.75]$
LAW-M ($R = 1$)	JPEG50	$86.12\ [81.93, 90.20]$	$0/128\ [0.00, 0.00]$
LAW-M ($R = 1$)	JPEG30	$81.84\ [76.99, 86.57]$	$0/128\ [0.00, 0.00]$
LAW-M ($R = 1$)	Resize256	$92.73\ [89.51, 95.33]$	$3/128\ [0.00, 7.81]$
LAW-M ($R = 1$)	Crop410	$50.13\ [49.61, 50.65]$	$0/128\ [0.00, 0.00]$

C.6 Recovery with SW numerical settings

This study evaluates ESM and SW under the full-precision and posterior-sampling settings described in Appendix A. All 256 generation requests pass the safety checker. Table 15 reports all three channels, and Table 16 gives the paired differences from 5000 crossed-bootstrap draws of the eight keys and sixteen prompts.

The JPEG50 BA advantage is 6.86 points, with a 95% interval of $[3.66, 10.58]$; the exact-recovery advantage is 64.84 points, with interval $[46.09, 80.47]$. Under JPEG70, ESM recovers 121/128 messages and SW 47/128. These paired results retain the recovery gap under the evaluated native numerical conventions. The earlier common-pipeline results remain a separate comparison because precision, posterior handling and SW sampling conventions change together.

Table 12: Paper-based LAW reconstruction recovery on SD2.1-base. Delivered BA and interval bounds are percentages; exact recovery is shown as count/128. Brackets give pointwise 95% crossed-bootstrap intervals.

Method	Channel	BA (%) [95% CI]	Exact/128 [rate CI (%)]
SD2.1-base			
LAW ($R = 7$)	Clean	97.00 [95.56, 98.22]	11/128 [0.78, 20.31]
LAW ($R = 7$)	JPEG50	84.70 [80.24, 88.88]	0/128 [0.00, 0.00]
LAW ($R = 7$)	JPEG30	79.45 [74.59, 84.35]	0/128 [0.00, 0.00]
LAW ($R = 7$)	Resize256	92.90 [90.13, 95.18]	4/128 [0.00, 10.94]
LAW ($R = 7$)	Crop410	50.20 [49.63, 50.73]	0/128 [0.00, 0.00]
LAW-M ($R = 1$)	Clean	97.54 [96.48, 98.49]	9/128 [0.00, 18.75]
LAW-M ($R = 1$)	JPEG50	84.79 [80.58, 88.82]	2/128 [0.00, 7.03]
LAW-M ($R = 1$)	JPEG30	79.95 [75.28, 84.57]	0/128 [0.00, 0.00]
LAW-M ($R = 1$)	Resize256	93.49 [91.28, 95.49]	0/128 [0.00, 0.00]
LAW-M ($R = 1$)	Crop410	49.83 [49.16, 50.50]	0/128 [0.00, 0.00]

Table 13: Secondary ESM-minus-LAW contrasts on SD1.5. Descriptor-matched cross-execution comparisons retain 128 requests per arm. Differences and pointwise 95% crossed-bootstrap intervals are in percentage points.

Comparator	Channel	Δ BA [95% CI]	Δ exact [95% CI]
SD1.5			
LAW ($R = 7$)	Clean	+3.70 [+2.03, +6.09]	+93.75 [+83.59, +100.00]
LAW ($R = 7$)	JPEG50	+16.01 [+11.47, +20.79]	+89.06 [+76.56, +97.66]
LAW ($R = 7$)	JPEG30	+21.03 [+15.72, +26.49]	+75.78 [+59.38, +89.84]
LAW ($R = 7$)	Resize256	+8.44 [+5.44, +12.10]	+96.88 [+89.84, +100.00]
LAW ($R = 7$)	Crop410	+1.22 [+0.30, +2.12]	+0.00 [+0.00, +0.00]
LAW-M ($R = 1$)	Clean	+2.80 [+1.49, +4.86]	+92.19 [+81.25, +99.22]
LAW-M ($R = 1$)	JPEG50	+13.70 [+9.76, +17.80]	+89.06 [+76.56, +97.66]
LAW-M ($R = 1$)	JPEG30	+17.76 [+13.26, +22.40]	+75.78 [+59.38, +89.84]
LAW-M ($R = 1$)	Resize256	+7.26 [+4.67, +10.54]	+95.31 [+86.72, +100.00]
LAW-M ($R = 1$)	Crop410	+1.16 [+0.29, +1.98]	+0.00 [+0.00, +0.00]

Table 14: Secondary ESM-minus-LAW contrasts on SD2.1-base. Descriptor-matched cross-execution comparisons retain 128 requests per arm. Differences and pointwise 95% crossed-bootstrap intervals are in percentage points.

Comparator	Channel	Δ BA [95% CI]	Δ exact [95% CI]
SD2.1-base			
LAW ($R = 7$)	Clean	+3.00 [+1.75, +4.58]	+91.41 [+79.69, +99.22]
LAW ($R = 7$)	JPEG50	+15.17 [+10.98, +19.53]	+85.94 [+72.66, +96.09]
LAW ($R = 7$)	JPEG30	+20.12 [+15.36, +24.74]	+76.56 [+60.16, +90.62]
LAW ($R = 7$)	Resize256	+7.10 [+4.75, +10.04]	+96.88 [+89.06, +100.00]
LAW ($R = 7$)	Crop410	+1.29 [+0.43, +2.15]	+0.00 [+0.00, +0.00]
LAW-M ($R = 1$)	Clean	+2.46 [+1.51, +3.54]	+92.97 [+81.25, +100.00]
LAW-M ($R = 1$)	JPEG50	+15.08 [+10.99, +19.22]	+84.38 [+70.31, +95.31]
LAW-M ($R = 1$)	JPEG30	+19.62 [+15.10, +23.99]	+76.56 [+60.16, +90.62]
LAW-M ($R = 1$)	Resize256	+6.51 [+4.48, +8.77]	+100.00 [+100.00, +100.00]
LAW-M ($R = 1$)	Crop410	+1.65 [+0.70, +2.58]	+0.00 [+0.00, +0.00]

C.7 Expanded amplitude and rotation ablation

The expanded study crosses all eight test keys and sixteen authored prompts on SD1.5. HN, rotated HN and ESM are regenerated together for each case and evaluated under clean and JPEG50 channels.

Table 15: **Recovery with SW numerical settings.** Both methods use full-precision SD1.5 and sampled VAE posterior inversion on 128 paired requests. Cells show BA (%) / exact messages out of 128; JPEG70 is included alongside the main JPEG50 condition.

Method	Clean	JPEG50	JPEG70
SW	99.63 / 109	92.96 / 32	95.99 / 47
ESM	100.00 / 128	99.82 / 115	99.92 / 121

Table 16: Paired ESM minus SW differences under the corresponding native numerical settings. Values are percentage points with 95% crossed-bootstrap intervals.

Channel	Δ BA [CI]	Δ exact [CI]
Clean	+0.37 [+0.02, +1.24]	+14.84 [+3.91, +29.69]
JPEG50	+6.86 [+3.66, +10.58]	+64.84 [+46.09, +80.47]
JPEG70	+3.93 [+1.84, +6.72]	+57.81 [+37.50, +77.34]

All 384 requests pass the safety checker. Matching the underlying Gaussian draw fixes block energy, while matching the frame between rotated HN and ESM isolates amplitude allocation through the complete generation–inversion pipeline.

Table 17: Paired JPEG50 differences in the 128-request amplitude study. Values are percentage points with 95% crossed-bootstrap intervals from 5000 key–prompt resamples.

Contrast	Δ BA [CI]	Δ exact [CI]
ESM minus HN	$+0.629$ $[+0.137, +1.230]$	$+39.062$ $[+22.656, +56.250]$
ESM minus Rotated HN	$+0.317$ $[-0.299, +0.780]$	$+24.219$ $[+10.938, +38.281]$
Rotated HN minus HN	$+0.311$ $[-0.142, +0.868]$	$+14.844$ $[-2.344, +34.375]$

The JPEG50 exact-message advantage of ESM over rotated HN is positive in every key group, ranging from 12.5 to 43.75 points. This gives a 24.22-point aggregate advantage with interval $[10.94, 38.28]$. The earlier 32-pair experiment had estimated 12.5 points with interval $[-6.25, 37.50]$; its smaller, overlapping set is retained separately in Appendix C.14. These intervals overlap, and the expanded study supplies the more precise complete-message comparison.

This execution has 64 exact HN recoveries under JPEG50, compared with 69 in the original main comparison; both have 114 for ESM. The request descriptors and all source latents match exactly across the two executions; differences arise during the subsequent generation–inversion runs. Their specific numerical cause has not been isolated. We report within-execution paired contrasts and do not pool the repeated requests. The full-pipeline HN contrast and the frame-matched amplitude contrast estimate different changes.

C.8 Finite-sampler and frame-reuse diagnostics

We compare ESM source noise with Gaussian noise and a synthetic shared-frame construction. Each construction has 24 samples for each of sixteen keys. The classifier is fitted on eight keys and evaluated on the other eight, giving 192 test samples per arm. Its 50 image-level features describe coordinate moments, normalized block moments, radii, pairwise Gram structure and fixed random projections. A diagonal linear discriminant uses a fixed ridge of 0.001. Feature normalization and the Gaussian-kernel bandwidth are determined from training data. The AUC interval resamples the eight held-out keys 2000 times; the MMD test enumerates all 256 paired key-group label swaps.

The shared-frame control has classifier and Gram-grid AUC 1.000. ESM has classifier AUC 0.480, with interval $[0.460, 0.501]$, and Gram-grid AUC 0.525 (Table 18). Its mean coordinate variance is 1.000629 and excess kurtosis is -0.001574, compared with 1.000631 and 0.000832 for Gaussian noise. The biased feature-space MMD squared is 0.002420 for ESM ($p = 0.992$) and 0.142862 for the shared-frame control ($p = 0.0078$). These are diagnostics of the specified finite sampler and feature representation, rather than a measured distance between the full latent distributions.

Table 18: **Testing shared-frame structure in the source noise.** Each arm is compared with Gaussian noise on eight held-out keys (192 samples per arm). Chance AUC is 0.5; brackets give a 95% key-bootstrap interval. MMD uses a grouped permutation test on fixed statistical features.

Construction	Classifier AUC [CI]	Gram AUC	MMD p
Synthetic shared frame	1.000 [1.000, 1.000]	1.000	0.008
ESM	0.480 [0.460, 0.501]	0.525	0.992

Within-key frame learning. We also apply the fixed fourth-order frame learner from Appendix F. On three held-out keys, ESM AUCs are 0.425, 0.596 and 0.436. However, the synthetic shared-frame control yields 0.483, 0.541 and 0.467, all below the prespecified sensitivity threshold of 0.9. This learner therefore gives no interpretable negative result for ESM in that experiment.

An exploratory follow-up applies the same learner to actual SW source noise, retaining the original sample descriptors, 128/32 train/test split per key, optimizer seed and 400-iteration budget. Its held-out AUC is 1.000 for each of the three keys. This establishes sensitivity to the actual SW construction in these conditions; it is separate from the failed synthetic-control experiment. Both outcomes are retained because sensitivity depends on the control construction and the structures available to the learner.

Diagnostics after image inversion. We repeat the statistical and frame-learning diagnostics on clean-image inversions for both backbones, comparing ESM, HN and SW with matched Gaussian controls. All 128 cases are jointly safe. The statistical classifier uses four key groups for fitting and four for testing. The within-key learner uses eight training and eight test prompts for each key, keeping its optimization rule fixed.

Table 19: Diagnostics on clean-image inversions. Each arm is compared with Gaussian A, except the A–B negative control. Classifier fitting and testing use disjoint sets of four keys. No SW control reaches the fixed 0.9 AUC sensitivity threshold.

Model	Contrast	Classifier AUC [95% CI]	Gram AUC	MMD p
SD1.5	ESM–A	0.536 [0.503, 0.562]	0.447	0.375
SD1.5	SW–A	0.517 [0.446, 0.598]	0.508	0.875
SD1.5	HN–A	0.526 [0.466, 0.593]	0.522	0.250
SD1.5	A–B	0.435 [0.375, 0.535]	0.605	0.875
SD2.1-base	ESM–A	0.500 [0.402, 0.586]	0.429	0.875
SD2.1-base	SW–A	0.560 [0.476, 0.655]	0.434	0.375
SD2.1-base	HN–A	0.464 [0.420, 0.510]	0.455	0.375
SD2.1-base	A–B	0.485 [0.418, 0.549]	0.568	0.125

Table 20: Within-key fourth-order learner on clean-image inversions. Each row retains all eight held-out key-specific AUCs, using eight training and eight test prompts per key. The SW controls do not reach the fixed sensitivity threshold.

Model	Contrast	Key 0	1	2	3	4	5	6	7
SD1.5	ESM–A	0.266	0.391	0.469	0.531	0.422	0.578	0.156	0.469
SD1.5	SW–A	0.562	0.641	0.453	0.531	0.625	0.688	0.594	0.547
SD1.5	A–B	0.438	0.641	0.609	0.234	0.469	0.203	0.562	0.578
SD2.1-base	ESM–A	0.344	0.484	0.531	0.438	0.609	0.609	0.750	0.625
SD2.1-base	SW–A	0.234	0.719	0.516	0.562	0.609	0.734	0.594	0.562
SD2.1-base	A–B	0.547	0.484	0.422	0.516	0.453	0.734	0.297	0.562

On SD1.5, the ESM classifier has AUC 0.536 with key-bootstrap interval $[0.503, 0.562]$, a small above-chance response in this test. On SD2.1-base it has AUC 0.500 with interval $[0.402, 0.586]$. All three inversion diagnostics fail to attain the 0.9 sensitivity threshold on their SW control. The within-key training sets are also much smaller than in the source-noise study. We therefore report these results as an inversion-stage diagnostic with unsuccessful control calibration, separately from

the source-noise evidence. With only four held-out keys, the grouped MMD test has sixteen label swaps and a minimum two-sided p-value of 0.125.

C.9 Image-quality metrics and controls

For each backbone, ESM and the two ordinary-Gaussian arms share the eight-key by sixteen-prompt request grid. Gaussian A uses the draw whose block norms supplied ESM's radii; Gaussian B uses a separate fixed random draw. ESM images are the original main-study outputs. All 128 triplets per model pass the safety checker. This paired design measures text alignment and output-distribution changes while the A–B comparison shows variation between two unwatermarked samples.

CLIP alignment (Radford et al., 2021) is the cosine between full CLIP ViT-B/32 image and text projections, using the actual prompt without an added prefix. Image diversity averages pairwise CLIP cosine distances between the eight outputs for each prompt, then averages the sixteen prompt values. FID (Heusel et al., 2017) uses the 2048-dimensional Inception pool-3 features from the TensorFlow-compatible FID network. These are empirical distances to generated Gaussian reference images, with 128 samples per arm, rather than FID against a real-image dataset. Their finite-sample bias is substantial, so the matched A–B distance is shown alongside the ESM–A distance.

KID (Bińkowski et al., 2018) uses the polynomial kernel $(x^\top y/2048 + 1)^3$ on Inception features. We use a paired off-diagonal U-statistic, excluding corresponding-case terms in all kernel sums. CLIP distribution diagnostics use polynomial $(x^\top y + 1)^3$ and Gaussian kernels on unit image embeddings. Kernel U-statistics may be negative. Pointwise intervals for alignment and kernel statistics use 5000 crossed key–prompt resamples; the descriptive FID intervals use 250 such draws.

The ESM-minus-Gaussian-A CLIP alignment difference is -0.0006 on SD1.5, with a 95% interval of $[-0.0082, 0.0062]$, and $+0.0006$ on SD2.1-base, with interval $[-0.0068, 0.0086]$.

Table 21: **Image quality and variation.** Each arm has 128 images per model. CLIP is text–image cosine similarity. FID and paired KID compare with generated Gaussian A; Gaussian B is an independent-noise control. Diversity is mean within-prompt CLIP distance.

Input noise	CLIP ↑	FID	KID $\times 10^3$	Diversity
SD1.5				
Gaussian A	0.3411	–	–	0.1088
Gaussian B	0.3424	113.12	0.502	0.0928
ESM	0.3405	110.54	-0.034	0.1046
SD2.1-base				
Gaussian A	0.3414	–	–	0.0915
Gaussian B	0.3403	102.59	-0.153	0.0941
ESM	0.3420	102.67	-0.277	0.0904

Image classifiers use standardized CLIP embeddings and a fixed regularized logistic model ($C = 1$), fitted on four key groups and evaluated on the remaining four. On SD2.1-base, ESM–A gives AUC 0.582 with interval $[0.551, 0.622]$, while the Gaussian A–B negative control gives 0.489 $[0.410, 0.544]$. On SD1.5, the corresponding values are 0.458 $[0.376, 0.527]$ and 0.553 $[0.480, 0.647]$. The SD2.1-base result identifies a feature-distribution difference in the evaluated images. Its classifier score measures separation, while the paired CLIP alignment contrast measures the direction of change in prompt similarity; the latter has an interval spanning zero. ESM and Gaussian images also come from separate generation executions, so isolating the source of the feature shift requires a contemporaneous comparison.

C.10 Diffusion regeneration

We regenerate SD1.5 outputs from eight keys and four fixed prompts, giving 32 matched HN/ESM pairs. Image-to-image generation uses an empty prompt, guidance 1 and 50 nominal steps at strengths 0.1 and 0.3. VAE posterior sampling and generation randomness are paired across the methods within each case and strength. The attacked images are then decoded with the common empty-prompt DDIM inversion pipeline. All 128 attacked images are delivered without safety erasure.

Table 22: Feature-space distribution diagnostics. All estimates and intervals are multiplied by 10^3. Paired off-diagonal U-statistic estimates may be negative; corresponding-case pairs are excluded. Intervals use 5000 crossed key–prompt resamples.

Model	Contrast	Statistic	Estimate	95% interval
SD1.5	ESM–A	Inception KID	-0.034	[-2.611, +2.638]
SD1.5	ESM–A	CLIP polynomial MMD	-2.634	[-19.635, +15.840]
SD1.5	ESM–A	CLIP RBF MMD	-0.268	[-2.014, +1.639]
SD1.5	A–B	Inception KID	+0.502	[-2.133, +4.686]
SD1.5	A–B	CLIP polynomial MMD	+2.001	[-12.964, +19.163]
SD1.5	A–B	CLIP RBF MMD	+0.202	[-1.339, +1.949]
SD2.1-base	ESM–A	Inception KID	-0.277	[-2.752, +2.354]
SD2.1-base	ESM–A	CLIP polynomial MMD	+0.110	[-13.415, +15.408]
SD2.1-base	ESM–A	CLIP RBF MMD	+0.012	[-1.370, +1.570]
SD2.1-base	A–B	Inception KID	-0.153	[-2.769, +2.965]
SD2.1-base	A–B	CLIP polynomial MMD	-1.038	[-16.815, +13.394]
SD2.1-base	A–B	CLIP RBF MMD	-0.103	[-1.700, +1.371]

Table 23: Image-feature classifier diagnostics using CLIP embeddings. Fits use four key groups and evaluation uses the remaining four (64 images per arm). Chance AUC is 0.5.

Model	Contrast	AUC	95% interval
SD1.5	ESM–A	0.458	[0.376, 0.527]
SD1.5	A–B	0.553	[0.480, 0.647]
SD2.1-base	ESM–A	0.582	[0.551, 0.622]
SD2.1-base	A–B	0.489	[0.410, 0.544]

Table 24: **Recovery after diffusion regeneration.** SD1.5, 32 paired requests per strength. Recovery is BA (%) / exact messages out of 32. PSNR, SSIM and LPIPS compare each attacked image with its own source image.

Strength	Method	Recovery ↑	PSNR ↑	SSIM ↑	LPIPS ↓
0.1	HN	99.31 / 16	23.59	0.705	0.059
0.1	**ESM**	99.95 / 24	23.32	0.701	0.060
0.3	HN	94.76 / 0	20.56	0.556	0.113
0.3	**ESM**	98.09 / 3	20.37	0.566	0.112

Table 24 reports recovery together with the attack's image distortion. At strength 0.1, ESM minus HN improves BA by 0.641 percentage points, with a 95% crossed-bootstrap interval of $[0.036, 1.532]$, and exact recovery by 25.0 points, with interval $[-12.5, 62.5]$. At strength 0.3, the corresponding differences are 3.326 points $[1.312, 5.359]$ and 9.375 points $[0, 28.125]$. Each interval uses 5000 independent resamples of the eight key groups and four prompt groups, retaining method pairing. Paired differences in PSNR, SSIM and LPIPS all have intervals containing zero. The image metrics measure distortion introduced by the attack relative to each method's original output. PSNR uses RGB intensities in $[0, 1]$. SSIM uses an 11×11 Gaussian window with standard deviation 1.5 and population moments, averaging the valid region over RGB channels (Wang et al., 2004). LPIPS uses the SqueezeNet backbone and version-0.1 perceptual weights (Zhang et al., 2018).

C.11 Matched modulation benchmark including HN

The updated benchmark measures HN, Householder ESM and dense-QR ESM on the same Intel Xeon Gold 6548Y+ CPU, using two numerical-library threads and float64 arrays. Each draw uses the same 512-bit payload, Gaussian amplitude draw, 128 blocks and block width 128. Two warmups precede twelve timed draws per method; the order cycles so that each method occupies each position

four times. Complete encoding and decoding each regenerate their mask or frame. Input validation, serialization, diffusion generation and inversion lie outside this modulation timing scope.

Table 25: **Cost of modulation and frame regeneration.** Medians of twelve CPU draws after two warmups, using two threads. Times include regenerating the mask or frame; diffusion is excluded. Arrays measure temporary frame/mask storage.

Method	Encode (ms)	Decode (ms)	Arrays (MiB)
HN	0.332	0.069	0.031
ESM, dense QR	190.683	191.740	16.000
ESM, Householder	49.812	49.294	8.188

Table 25 includes the inexpensive HN reference. Householder reconstruction reduces ESM encoding and decoding time by factors of 3.83 and 3.89 relative to dense QR. The frame construction costs roughly 50 ms per operation beyond HN. Applying a frame that has already been generated is a different operation: QR forward/inverse application takes 0.747/0.894 ms, versus 5.280/4.990 ms for Householder factors. Reconstruction accounts for the complete-operation speedup.

The temporary mask/frame representations occupy 0.031, 8.188 and 16.000 MiB for HN, Householder and QR, respectively. Separate processes on the same host reach peak resident sizes of 47.55, 78.72 and 111.13 MiB after the benchmark. These process measurements include the interpreter and libraries. All methods retain a 32-byte global key and a 12-byte per-image nonce; no frame is cached between encoding and decoding. The older timing study below is preserved as a separate measurement on different hardware.

C.12 Original-study native and auxiliary decoder cells

GS and SW retain their native hard rules; HN and ESM use soft sums. Rows labeled native repeat the corresponding hard or soft result to make each method's main comparison easy to locate. The auxiliary rules use each method's own layout. Delivered accuracy includes erasures as zeros, whereas conditional accuracy averages safe requests; the columns coincide because no erasures occurred.

Table 26: SD1.5, Clean: complete 512-bit decoder results. BA is in %; delivered BA includes safety erasures, conditional BA uses safe images only. Exact and erased are counts out of 128 requests.

Method	Decoder	Delivered BA	Conditional BA	Exact	Erased
HN	native	99.985	99.985	124	0
HN	hard	99.933	99.933	110	0
HN	soft	99.985	99.985	124	0
GS	native	99.956	99.956	117	0
GS	hard	99.956	99.956	117	0
GS	soft	99.995	99.995	126	0
SW	native	99.672	99.672	111	0
SW	hard	99.672	99.672	111	0
SW	soft	99.776	99.776	115	0
ESM	native	100.000	100.000	128	0
ESM	hard	99.997	99.997	126	0
ESM	soft	100.000	100.000	128	0

C.13 Paired JPEG50 contrasts and key groups

The following pointwise 95% intervals resample key and prompt groups independently while keeping methods paired. ESM–HN is the primary comparison within each backbone; GS and SW comparisons are secondary. Per-key means show the variation behind the aggregate result. Amplitude-ablation intervals use their own four-key by eight-prompt crossing.

Table 27: SD1.5, JPEG50: complete 512-bit decoder results. BA is in %; delivered BA includes safety erasures, conditional BA uses safe images only. Exact and erased are counts out of 128 requests.

Method	Decoder	Delivered BA	Conditional BA	Exact	Erased
HN	native	99.194	99.194	69	0
HN	hard	98.354	98.354	41	0
HN	soft	99.194	99.194	69	0
GS	native	98.657	98.657	46	0
GS	hard	98.657	98.657	46	0
GS	soft	99.501	99.501	68	0
SW	native	93.597	93.597	35	0
SW	hard	93.597	93.597	35	0
SW	soft	94.946	94.946	37	0
ESM	native	99.821	99.821	114	0
ESM	hard	99.594	99.594	95	0
ESM	soft	99.821	99.821	114	0

Table 28: SD1.5, JPEG30: complete 512-bit decoder results. BA is in %; delivered BA includes safety erasures, conditional BA uses safe images only. Exact and erased are counts out of 128 requests.

Method	Decoder	Delivered BA	Conditional BA	Exact	Erased
HN	native	98.271	98.271	49	0
HN	hard	96.942	96.942	28	0
HN	soft	98.271	98.271	49	0
GS	native	97.394	97.394	34	0
GS	hard	97.394	97.394	34	0
GS	soft	98.785	98.785	52	0
SW	native	88.834	88.834	19	0
SW	hard	88.834	88.834	19	0
SW	soft	90.753	90.753	24	0
ESM	native	99.599	99.599	97	0
ESM	hard	99.046	99.046	69	0
ESM	soft	99.599	99.599	97	0

Table 29: SD1.5, Resize: complete 512-bit decoder results. BA is in %; delivered BA includes safety erasures, conditional BA uses safe images only. Exact and erased are counts out of 128 requests.

Method	Decoder	Delivered BA	Conditional BA	Exact	Erased
HN	native	99.881	99.881	106	0
HN	hard	99.644	99.644	77	0
HN	soft	99.881	99.881	106	0
GS	native	99.713	99.713	85	0
GS	hard	99.713	99.713	85	0
GS	soft	99.918	99.918	113	0
SW	native	98.125	98.125	64	0
SW	hard	98.125	98.125	64	0
SW	soft	98.535	98.535	82	0
ESM	native	99.989	99.989	125	0
ESM	hard	99.933	99.933	121	0
ESM	soft	99.989	99.989	125	0

C.14 Amplitude-control decoder details

The 32-pair rotated-HN/ESM comparison holds the frame and block energy fixed while changing coordinate magnitudes. The tables report both auxiliary and native decoders on all preselected

Table 30: SD1.5, Crop: complete 512-bit decoder results. BA is in %; delivered BA includes safety erasures, conditional BA uses safe images only. Exact and erased are counts out of 128 requests.

Method	Decoder	Delivered BA	Conditional BA	Exact	Erased
HN	native	51.157	51.157	0	0
HN	hard	50.917	50.917	0	0
HN	soft	51.157	51.157	0	0
GS	native	50.642	50.642	0	0
GS	hard	50.642	50.642	0	0
GS	soft	50.728	50.728	0	0
SW	native	49.968	49.968	0	0
SW	hard	49.968	49.968	0	0
SW	soft	49.783	49.783	0	0
ESM	native	51.294	51.294	0	0
ESM	hard	50.922	50.922	0	0
ESM	soft	51.294	51.294	0	0

Table 31: SD2.1-base, Clean: complete 512-bit decoder results. BA is in %; delivered BA includes safety erasures, conditional BA uses safe images only. Exact and erased are counts out of 128 requests.

Method	Decoder	Delivered BA	Conditional BA	Exact	Erased
HN	native	99.985	99.985	125	0
HN	hard	99.945	99.945	115	0
HN	soft	99.985	99.985	125	0
GS	native	99.980	99.980	119	0
GS	hard	99.980	99.980	119	0
GS	soft	99.997	99.997	126	0
SW	native	99.915	99.915	115	0
SW	hard	99.915	99.915	115	0
SW	soft	99.953	99.953	121	0
ESM	native	100.000	100.000	128	0
ESM	hard	100.000	100.000	128	0
ESM	soft	100.000	100.000	128	0

cases. The JPEG50 BA interval touches zero and the exact-message interval spans zero, showing the uncertainty around the positive point estimates.

C.15 Development-calibrated likelihood decoding

The working observation model is $U = gSA + \varepsilon$, with $\varepsilon \sim N(0, \sigma^2)$, $A = |N(0,1)|$ for HN and $A = \chi_{128}/\sqrt{128}$ for ESM. Each of the twenty model–method–channel cells fits $g = \sum xy / \sum x^2$ and residual-RMS σ on the safe pairs among eight DEV requests, using a zero intercept. TEST decoding sums coordinate marginal log-likelihood ratios across the declared repeats and thresholds at zero. Its inputs are the recovered coordinates and fitted parameters; calibration supplies a scalar approximation to the image channel.

For HN the marginal LLR has an analytic Gaussian-CDF expression. For ESM, numerical quadrature evaluates the ideal radius marginal. Parameters are estimated on development data and kept fixed for all 128 test images in each cell. Reported changes are descriptive paired comparisons.

C.16 Conditional geometric diagnostic

Every source/recovered pair determines the longitudinal and transverse terms of the conditional joint law. Each of 512 draws simulates a complete vector in the all-ones orthogonal complement for every block, adds the longitudinal component, and sums repetitions with the original four-group layout.

Table 32: SD2.1-base, JPEG50: complete 512-bit decoder results. BA is in %; delivered BA includes safety erasures, conditional BA uses safe images only. Exact and erased are counts out of 128 requests.

Method	Decoder	Delivered BA	Conditional BA	Exact	Erased
HN	native	99.226	99.226	74	0
HN	hard	98.491	98.491	47	0
HN	soft	99.226	99.226	74	0
GS	native	98.886	98.886	45	0
GS	hard	98.886	98.886	45	0
GS	soft	99.574	99.574	79	0
SW	native	94.228	94.228	35	0
SW	hard	94.228	94.228	35	0
SW	soft	95.299	95.299	40	0
ESM	native	99.873	99.873	110	0
ESM	hard	99.620	99.620	95	0
ESM	soft	99.873	99.873	110	0

Table 33: SD2.1-base, JPEG30: complete 512-bit decoder results. BA is in %; delivered BA includes safety erasures, conditional BA uses safe images only. Exact and erased are counts out of 128 requests.

Method	Decoder	Delivered BA	Conditional BA	Exact	Erased
HN	native	98.502	98.502	50	0
HN	hard	97.202	97.202	30	0
HN	soft	98.502	98.502	50	0
GS	native	97.542	97.542	33	0
GS	hard	97.542	97.542	33	0
GS	soft	98.924	98.924	53	0
SW	native	89.304	89.304	22	0
SW	hard	89.304	89.304	22	0
SW	soft	91.370	91.370	25	0
ESM	native	99.568	99.568	98	0
ESM	hard	98.976	98.976	81	0
ESM	soft	99.568	99.568	98	0

Thus within-block bit dependence is retained, while distinct blocks use independent residual draws. No channel parameter is fitted. The analysis uses all 256 ESM source latents and 1280 inversions.

The table separates four quantities: aggregate observed/predicted BER; observed exact rate versus mean conditional exact probability; the mean per-image absolute BER discrepancy; and mean Monte Carlo standard error (MC SE) of each image's estimated conditional BER. The last quantity measures finite simulation error, not sampling uncertainty over keys/prompts. Coverage is the descriptive fraction of images whose actual error count lies between the simulated 5th and 95th percentiles. Discrete near-zero error counts can make this central interval conservative; its coverage is not a formal calibration test.

The largest aggregate exact-rate discrepancy is 3.04 percentage points (SD1.5 JPEG30, observed 75.78% versus predicted 72.74%). The crop has the largest per-image BER discrepancies, about 1.12 and 1.04 points, and central error-count interval coverages 89.06% and 89.84%. These quantities show where the geometric prediction is most accurate and where individual images depart most from its conditional mean.

C.17 Regeneration cost and working storage

The compact comparison from the main-text analysis is reported here together with the complete timing scopes and memory measurements. The timing experiment uses an Intel Xeon Gold 6248

Table 34: SD2.1-base, Resize: complete 512-bit decoder results. BA is in %; delivered BA includes safety erasures, conditional BA uses safe images only. Exact and erased are counts out of 128 requests.

Method	Decoder	Delivered BA	Conditional BA	Exact	Erased
HN	native	99.913	99.913	109	0
HN	hard	99.724	99.724	83	0
HN	soft	99.913	99.913	109	0
GS	native	99.844	99.844	85	0
GS	hard	99.844	99.844	85	0
GS	soft	99.982	99.982	120	0
SW	native	99.216	99.216	68	0
SW	hard	99.216	99.216	68	0
SW	soft	99.464	99.464	84	0
ESM	native	100.000	100.000	128	0
ESM	hard	99.992	99.992	125	0
ESM	soft	100.000	100.000	128	0

Table 35: SD2.1-base, Crop: complete 512-bit decoder results. BA is in %; delivered BA includes safety erasures, conditional BA uses safe images only. Exact and erased are counts out of 128 requests.

Method	Decoder	Delivered BA	Conditional BA	Exact	Erased
HN	native	50.757	50.757	0	0
HN	hard	50.662	50.662	0	0
HN	soft	50.757	50.757	0	0
GS	native	50.818	50.818	0	0
GS	hard	50.818	50.818	0	0
GS	soft	50.781	50.781	0	0
SW	native	50.157	50.157	0	0
SW	hard	50.157	50.157	0	0
SW	soft	50.131	50.131	0	0
ESM	native	51.486	51.486	0	0
ESM	hard	51.166	51.166	0	0
ESM	soft	51.486	51.486	0	0

Table 36: JPEG50 native contrasts, ESM minus comparator, in percentage points. Pointwise 95% crossed-bootstrap intervals use 5000 draws of keys and prompts with method pairing; no multiplicity adjustment. The HN contrast is primary; GS/SW contrasts are secondary. Exact denotes exact-message rate.

Model	Comparator	Δ BA	95% interval	Δ exact	95% interval
SD1.5	HN	+0.63	[+0.12, +1.25]	+35.16	[+17.97, +52.34]
SD1.5	GS	+1.16	[+0.44, +2.05]	+53.12	[+34.38, +71.09]
SD1.5	SW	+6.22	[+3.40, +9.37]	+61.72	[+43.75, +78.91]
SD2.1-base	HN	+0.65	[+0.09, +1.56]	+28.12	[+11.72, +46.88]
SD2.1-base	GS	+0.99	[+0.49, +1.59]	+50.78	[+34.38, +67.19]
SD2.1-base	SW	+5.65	[+3.26, +8.31]	+58.59	[+39.06, +77.34]

at 2.50 GHz, float64, 128 blocks of width 128, and two numerical-library threads. Both arms use the same pseudorandom generator and inverse-normal sampler, 512-bit payload, dimensions, precision and paired input draws. Algorithm-specific streams and normal counts differ. Two warmups precede ten timed paired draws, with algorithm order alternating. Complete encode and decode each regenerate their frame; no matrix or factor is cached between them.

Table 37: JPEG50 per-key ESM minus HN native differences (percentage points), 16 paired prompt cases per key. These are differences only; no per-key absolute rates are inferred from the comparison report.

Key	SD1.5 BA	SD1.5 exact	SD2.1 BA	SD2.1 exact
0	+1.221	+31.250	-0.049	+31.250
1	+0.488	+43.750	+0.452	+50.000
2	+1.440	+43.750	+1.843	+43.750
3	+0.146	+31.250	+0.208	+37.500
4	+0.623	+25.000	+0.513	+25.000
5	+0.525	+25.000	+0.183	+18.750
6	+0.146	+31.250	+0.745	+0.000
7	+0.427	+50.000	+1.282	+18.750

Table 38: Standalone rotation ablation, JPEG50: ESM minus rotated HN in percentage points, 32 pairs. Pointwise 95% percentile crossed-bootstrap intervals use 5000 draws; four keys and eight prompts are resampled with method pairing. No multiplicity adjustment.

Metric	Difference	95% interval
BA	+0.250	[+0.00, +0.87]
Exact	+12.500	[-6.25, +37.50]

Table 39: **Paired improvements under JPEG50.** ESM minus comparator, in percentage points, with pointwise 95% crossed-bootstrap intervals from 5000 key–prompt draws. Shaded HN rows are the prespecified main comparisons; GS/SW comparisons are secondary. The amplitude ablation uses a separate set of 32 pairs.

Comparator	**Bit accuracy: Δ [95% CI]**	**Exact recovery: Δ [95% CI]**
SD1.5: main study, 128 pairs		
HN	$+0.63\ [+0.12, +1.25]$	$+35.16\ [+17.97, +52.34]$
GS	$+1.16\ [+0.44, +2.05]$	$+53.12\ [+34.38, +71.09]$
SW	$+6.22\ [+3.40, +9.37]$	$+61.72\ [+43.75, +78.91]$
SD2.1-base: main study, 128 pairs		
HN	$+0.65\ [+0.09, +1.56]$	$+28.12\ [+11.72, +46.88]$
GS	$+0.99\ [+0.49, +1.59]$	$+50.78\ [+34.38, +67.19]$
SW	$+5.65\ [+3.26, +8.31]$	$+58.59\ [+39.06, +77.34]$
SD1.5: amplitude ablation, 32 pairs		
Rotated HN	$+0.250\ [+0.00, +0.87]$	$+12.500\ [-6.25, +37.50]$

Table 40: SD1.5, Clean: complete 512-bit decoder results. BA is in %; delivered BA includes safety erasures, conditional BA uses safe images only. Exact and erased are counts out of 32 requests.

Method	Decoder	Delivered BA	Conditional BA	Exact	Erased
Rotated HN	native	99.969	99.969	30	0
Rotated HN	hard	99.890	99.890	29	0
Rotated HN	soft	99.969	99.969	30	0
ESM	native	100.000	100.000	32	0
ESM	hard	100.000	100.000	32	0
ESM	soft	100.000	100.000	32	0

Table 50 measures the implementation tradeoff. Seeded QR is faster when an already generated matrix is merely applied: forward/inverse medians are 0.854/1.082 ms, versus 6.917/6.682 ms for Householder factors. Regeneration reverses the complete-operation comparison. Householder encode/decode medians are 67.752/68.778 ms versus 245.789/244.985 ms for seeded QR. The approx-

Table 41: SD1.5, JPEG50: complete 512-bit decoder results. BA is in %; delivered BA includes safety erasures, conditional BA uses safe images only. Exact and erased are counts out of 32 requests.

Method	Decoder	Delivered BA	Conditional BA	Exact	Erased
Rotated HN	native	99.695	99.695	25	0
Rotated HN	hard	99.133	99.133	17	0
Rotated HN	soft	99.695	99.695	25	0
ESM	native	99.945	99.945	29	0
ESM	hard	99.811	99.811	24	0
ESM	soft	99.945	99.945	29	0

Table 42: SD1.5, JPEG30: complete 512-bit decoder results. BA is in %; delivered BA includes safety erasures, conditional BA uses safe images only. Exact and erased are counts out of 32 requests.

Method	Decoder	Delivered BA	Conditional BA	Exact	Erased
Rotated HN	native	99.329	99.329	22	0
Rotated HN	hard	98.383	98.383	14	0
Rotated HN	soft	99.329	99.329	22	0
ESM	native	99.860	99.860	28	0
ESM	hard	99.506	99.506	17	0
ESM	soft	99.860	99.860	28	0

Table 43: SD1.5, Resize: complete 512-bit decoder results. BA is in %; delivered BA includes safety erasures, conditional BA uses safe images only. Exact and erased are counts out of 32 requests.

Method	Decoder	Delivered BA	Conditional BA	Exact	Erased
Rotated HN	native	99.921	99.921	28	0
Rotated HN	hard	99.512	99.512	25	0
Rotated HN	soft	99.921	99.921	28	0
ESM	native	99.969	99.969	31	0
ESM	hard	99.902	99.902	29	0
ESM	soft	99.969	99.969	31	0

Table 44: SD1.5, Crop: complete 512-bit decoder results. BA is in %; delivered BA includes safety erasures, conditional BA uses safe images only. Exact and erased are counts out of 32 requests.

Method	Decoder	Delivered BA	Conditional BA	Exact	Erased
Rotated HN	native	51.196	51.196	0	0
Rotated HN	hard	50.909	50.909	0	0
Rotated HN	soft	51.196	51.196	0	0
ESM	native	51.794	51.794	0	0
ESM	hard	51.538	51.538	0	0
ESM	soft	51.794	51.794	0	0

imately 3.63/3.56 ratios concern these CPU modulation operations, not image generation, inversion or end-to-end throughput. Both constructions can regenerate from the same sizes of retained key/nonce metadata; QR has no mandatory dense per-image cache.

The frame arrays occupy 16.000 MiB for QR and 8.1875 MiB for the packed Householder representation. At completion of frame generation, the live array payloads total 48.125/16.251 MiB. These sums exclude other runtime allocations and are not measured peak memory.

Process memory is measured separately on an Intel Xeon 6767P, using one process per algorithm and two numerical-library threads. After two warmups and ten draws, process high-water RSS is 112.816 MiB for QR and 78.535 MiB for Householder (Table 51). These single measurements include

Table 45: SD1.5: native versus DEV-calibrated marginal-LLR decoding at 512 bits, evaluated on TEST. Each decoder cell is delivered BA (%) / exact-message count out of 128 requests. Δ BA is LLR minus native in percentage points; erased requests remain in all denominators. Changes are descriptive; the fitted scalar Gaussian channel and repetition independence are working assumptions.

Method	Channel	Native BA / exact	LLR BA / exact	Δ BA	Erased
HN	Clean	99.98 / 124	99.98 / 123	-0.002	0
HN	JPEG50	99.19 / 69	99.21 / 70	+0.017	0
HN	JPEG30	98.27 / 49	98.25 / 46	-0.018	0
HN	Resize	99.88 / 106	99.88 / 107	-0.005	0
HN	Crop	51.16 / 0	51.16 / 0	+0.000	0
ESM	Clean	100.00 / 128	100.00 / 128	+0.000	0
ESM	JPEG50	99.82 / 114	99.82 / 114	+0.000	0
ESM	JPEG30	99.60 / 97	99.60 / 97	+0.000	0
ESM	Resize	99.99 / 125	99.99 / 125	+0.000	0
ESM	Crop	51.29 / 0	51.29 / 0	+0.000	0

Table 46: SD1.5: calibration parameters fitted on DEV and held fixed for TEST. Each cell requests 8 cases (keys 8–9, prompts 16–19). Gain uses zero-intercept pooled OLS; σ is residual RMS with denominator equal to the number of coordinates. Only safe source/recovered DEV pairs enter the fit; erased count is 8 minus safe count. TEST results are not used for fitting.

Method	Channel	Gain	σ	Safe	Coordinates
HN	Clean	0.7761315	0.544434	8	131072
HN	JPEG50	0.6320644	0.7271454	8	131072
HN	JPEG30	0.5795025	0.7721664	8	131072
HN	Resize	0.7130946	0.6543105	8	131072
HN	Crop	0.004695757	0.9602159	8	131072
ESM	Clean	0.8058494	0.5063704	8	131072
ESM	JPEG50	0.6406917	0.7206821	8	131072
ESM	JPEG30	0.5858613	0.7669673	8	131072
ESM	Resize	0.7567162	0.6021881	8	131072
ESM	Crop	0.001633149	0.9668068	8	131072

Table 47: SD2.1-base: native versus DEV-calibrated marginal-LLR decoding at 512 bits, evaluated on TEST. Each decoder cell is delivered BA (%) / exact-message count out of 128 requests. Δ BA is LLR minus native in percentage points; erased requests remain in all denominators. Changes are descriptive; the fitted scalar Gaussian channel and repetition independence are working assumptions.

Method	Channel	Native BA / exact	LLR BA / exact	Δ BA	Erased
HN	Clean	99.98 / 125	99.98 / 123	-0.003	0
HN	JPEG50	99.23 / 74	99.21 / 75	-0.012	0
HN	JPEG30	98.50 / 50	98.49 / 49	-0.009	0
HN	Resize	99.91 / 109	99.91 / 109	-0.008	0
HN	Crop	50.76 / 0	50.76 / 0	+0.000	0
ESM	Clean	100.00 / 128	100.00 / 128	+0.000	0
ESM	JPEG50	99.87 / 110	99.87 / 110	+0.000	0
ESM	JPEG30	99.57 / 98	99.57 / 98	+0.000	0
ESM	Resize	100.00 / 128	100.00 / 128	+0.000	0
ESM	Crop	51.49 / 0	51.49 / 0	+0.000	0

the interpreter, imported libraries, allocator and benchmark overhead; they are neither incremental allocations nor per-operation peaks. Memory and timing measurements use different CPU models.

Table 48: SD2.1-base: calibration parameters fitted on DEV and held fixed for TEST. Each cell requests 8 cases (keys 8–9, prompts 16–19). Gain uses zero-intercept pooled OLS; σ is residual RMS with denominator equal to the number of coordinates. Only safe source/recovered DEV pairs enter the fit; erased count is 8 minus safe count. TEST results are not used for fitting.

Method	Channel	Gain	σ	Safe	Coordinates
HN	Clean	0.8036214	0.5078715	8	131072
HN	JPEG50	0.5971911	0.7573661	8	131072
HN	JPEG30	0.5295819	0.807711	8	131072
HN	Resize	0.7350203	0.6213784	8	131072
HN	Crop	0.003542464	0.960854	8	131072
ESM	Clean	0.7913914	0.5240421	7	114688
ESM	JPEG50	0.5750333	0.7746939	7	114688
ESM	JPEG30	0.5185602	0.8158494	7	114688
ESM	Resize	0.6805375	0.6832195	7	114688
ESM	Crop	0.007345303	0.9598217	7	114688

Table 49: Complete geometric diagnostic. BER and exact columns give observed / predicted values in percent. MAE and mean MC SE are percentage points; coverage is in percent. Each row uses 128 images, each with 512 conditional draws. Reported zeros are empirical values at the displayed precision.

Channel	BER (obs/pred)	Exact (obs/pred)	MAE	MC SE	Coverage
SD1.5					
Clean	0.0000 / 0.0002	100.00 / 99.91	0.0002	0.00005	100.00
JPEG50	0.1785 / 0.1792	89.06 / 89.04	0.0356	0.00224	97.66
JPEG30	0.4013 / 0.4183	75.78 / 72.74	0.0785	0.00529	99.22
Resize256	0.0107 / 0.0128	97.66 / 97.26	0.0063	0.00045	100.00
Crop410	48.7061 / 48.6093	0.00 / 0.00	1.1218	0.05882	89.06
SD2.1-base					
Clean	0.0000 / 0.0000	100.00 / 100.00	0.0000	0.00000	100.00
JPEG50	0.1266 / 0.1089	85.94 / 87.15	0.0490	0.00239	97.66
JPEG30	0.4318 / 0.4317	76.56 / 77.37	0.0815	0.00528	98.44
Resize256	0.0000 / 0.0005	100.00 / 99.76	0.0005	0.00009	100.00
Crop410	48.5138 / 48.5670	0.00 / 0.00	1.0356	0.05883	89.84

Table 50: CPU frame benchmark: medians of 10 paired timed draws after 2 warmups, float64, 128 blocks of width 128, 512-bit payload, two numerical-library threads. Encode/decode include frame regeneration. Frame MiB counts working frame arrays, not retained per-image metadata or peak memory. Both methods retain only a 12-byte nonce per image and a 32-byte global key, with no frame cached between encoding and decoding. Diffusion and inversion are outside these timings. Bold marks the lower measured value in this matched microbenchmark.

Frame	Generate (ms)	Encode (ms)	Decode (ms)	Frame (MiB)
Seeded QR	239.605	245.789	244.985	16.000
Householder	**64.580**	**67.752**	**68.778**	**8.188**

Table 51: Process high-water RSS on an Intel Xeon 6767P, measured after imports and after the benchmark workload. Each algorithm is measured once in a separate process. RSS includes runtime overhead and is not incremental allocation. Timing measurements use an Intel Xeon Gold 6248.

Frame	Post-import RSS (MiB)	Workload RSS (MiB)
Seeded QR	45.203	112.816
Householder	45.207	78.535

Table 52: Complete CPU resource records. Timing panels report milliseconds over 10 paired draws; sample SD uses denominator 9. Application scopes use an already generated frame, whereas complete encode/decode regenerate it. The memory panel reports array payloads rather than runtime overhead or peak memory; neither method caches a frame between encoding and decoding.

(a) Seeded QR timing scopes.

Scope	Median	Min	Max	Mean	Sample SD
Frame generation	239.604701	235.826156	243.524064	239.483747	2.969118
Forward application	0.853643	0.836903	0.885964	0.856024	0.017594
Inverse application	1.081997	1.010114	1.106063	1.067742	0.037140
Encode incl. regeneration	245.789001	243.261280	257.205490	246.909001	4.360269
Decode incl. regeneration	244.985224	242.679902	246.355409	244.839393	1.230543

(b) Householder timing scopes.

Scope	Median	Min	Max	Mean	Sample SD
Frame generation	64.579558	60.272846	70.703706	64.348831	3.455824
Forward application	6.916976	6.751585	7.801047	6.989537	0.296848
Inverse application	6.682446	6.405365	7.529334	6.708495	0.318151
Encode incl. regeneration	67.752178	67.174177	70.899939	68.233384	1.186121
Decode incl. regeneration	68.778147	66.481344	70.662085	68.439305	1.513116

(c) Array payloads (MiB); retained metadata is a 12-byte nonce per image and a 32-byte global key.

Frame	Frame arrays	Arrays at generation completion
Seeded QR	16.000000	48.125000
Householder	8.187500	16.250977

D HISTORICAL 1024-BIT NONCE STUDY

This appendix reports the earlier SD1.5 study at 1024 bits. Its results are separate from the two-backbone 512-bit study; references to the primary comparison below concern this historical protocol.

D.1 STUDY DESIGN

The protocol was specified after exploring soft decoding in the earlier 512-bit pilot and before observing the 1024-bit results. Four new public keys and sixteen new authored prompts yielded 64 paired cases, 128 generation requests and five channels per image. We used SD1.5 in fp16, 50-step DPM-family generation with guidance 7.5, VAE posterior mode and 50-step empty-prompt DDIM inversion with guidance 1. Channels were clean, JPEG50, JPEG30, Lanczos resize to 256^2 and back to 512^2, and centered 410^2 crop resized to 512^2 (about 64% area retained). Prompts, decoders and channel settings remained fixed.

The generation safety checker remained active. Flagged requests received zero delivered accuracy and were excluded from inversion and conditional safe-image accuracy. Complete-message counts used all 64 requests per arm. The primary comparison was the JPEG50 delivered soft-accuracy difference. Its 95% percentile interval independently resampled four key groups and sixteen prompt groups with 5000 crossed-bootstrap draws. Payloads and amplitude draws used public random seeds.

Recovery results. The study includes 128 safe images and 640 channel inversions. The primary JPEG50 soft-accuracy difference is 1.46 percentage points, with a 95% interval of [0.21, 3.27]. Complete 1024-bit payloads are recovered for 47/64 ESM images and 14/64 half-normal images.

Table 53: Complete 1024-bit study. HN: half-normal; ESM: nonce frames. Soft accuracy includes safety erasures as zero. Exact counts use all 64 requests. The prespecified primary channel is labeled in bold; numerical emphasis does not replace the paired interval reported in the text.

	Soft accuracy (%)			Exact messages	
Channel	HN	ESM	Δ (pp)	HN	ESM
Clean	99.95	100.00	+0.05	47/64	63/64
JPEG 50	97.80	99.25	+1.46	14/64	47/64
JPEG 30	96.13	98.62	+2.49	7/64	30/64
Resize 256	99.52	99.99	+0.47	26/64	59/64
Crop 410	50.76	50.55	-0.21	0/64	0/64

Safety failures are 0/64 for HN and 0/64 for ESM. Conditional JPEG50 soft accuracy is 97.80% and 99.25%, respectively, compared with hard accuracy of 95.95% and 98.50%.

Implementation cost. On a four-CPU host, ten benchmark draws average 412.1 ms for the earlier dense-QR encoder and 44.0 ms for the factorized encoder (9.36x ratio). The PRG and payload size also differ, while latent dimensions match, so this comparison does not isolate factorization cost. In the image study ESM averages 47.0 ms encoding and 45.1 ms decoding including regeneration. The earlier cached 2 ms decoder excluded regeneration and is therefore not directly comparable.

Recovery and computation. All four key-group JPEG50 soft differences were positive, ranging from 0.26 to 2.26 percentage points. JPEG30 gave a secondary mean difference of 2.49 points. Under cropping, ESM and HN had 50.55% and 50.76% soft accuracy, respectively, with zero complete messages in either arm. HN's mean JPEG50 encode/decode times were 0.55/0.31ms, versus about 47/45ms for ESM with regeneration. These measurements show the computation paid for the spherical construction. Factorization reduced the cost relative to dense ESM, whose float32 frame cache occupied 8 MiB in this implementation.

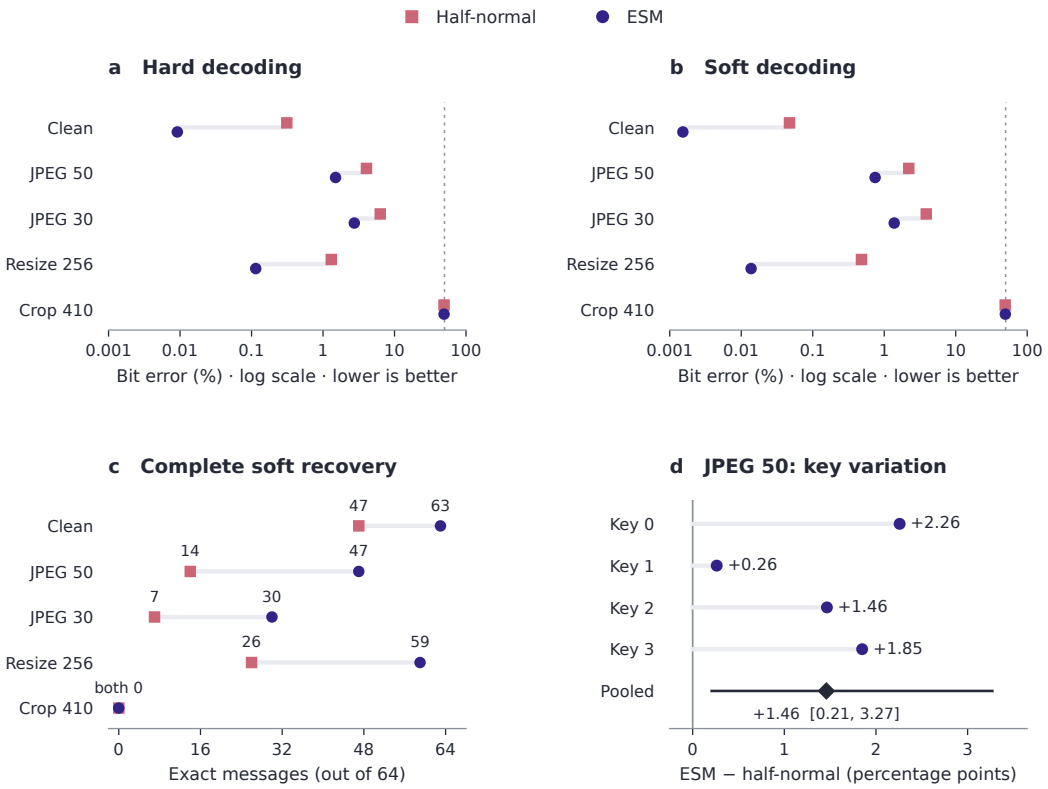


Figure 12: The complete nonce-study comparison. (a,b) Hard and soft bit error, 100% minus delivered accuracy, on a logarithmic axis; lower is better, and the reference line marks 50%. (c) Exact soft-message recovery out of 64 requests. (d) Each key's JPEG50 soft-accuracy difference and the pooled estimate with its 95% crossed-bootstrap interval. Key dots are descriptive means, not intervals; the pooled interval also resamples prompts. No best-key or best-channel selection.

D.2 Decoder ablation and variation across keys

The matched-soft comparison tests whether the gain persists when both methods retain reliability information. Under JPEG50, moving from hard to soft decisions raises the half-normal control from 95.95% to 97.80%, while ESM rises from 98.50% to 99.25%. The control benefits more in percentage-point terms, narrowing the gap to the primary 1.46-point difference.

Table 54: JPEG50 soft payload accuracy by key. Each row averages the same 16 prompts in both arms. The aggregate interval resamples both keys and prompts.

Test key	Half-normal (%)	ESM (%)	Δ (pp)
Key 0	97.455	99.713	+2.258
Key 1	98.669	98.932	+0.262
Key 2	97.876	99.341	+1.465
Key 3	97.180	99.030	+1.849

The JPEG50 difference is positive for all four tested keys. Key 1 has the strongest control mean and the smallest gap. The crossed bootstrap summarizes variation across these four keys and sixteen authored prompts.

Exact-message recovery distinguishes small bit-error rates from reliable identifier recovery. In the clean condition, ESM's rounded soft mean is 100.00%, but only 63 of 64 messages are exact. Under JPEG50, its 99.25% mean accompanies 47 exact messages, compared with 14 for the control.

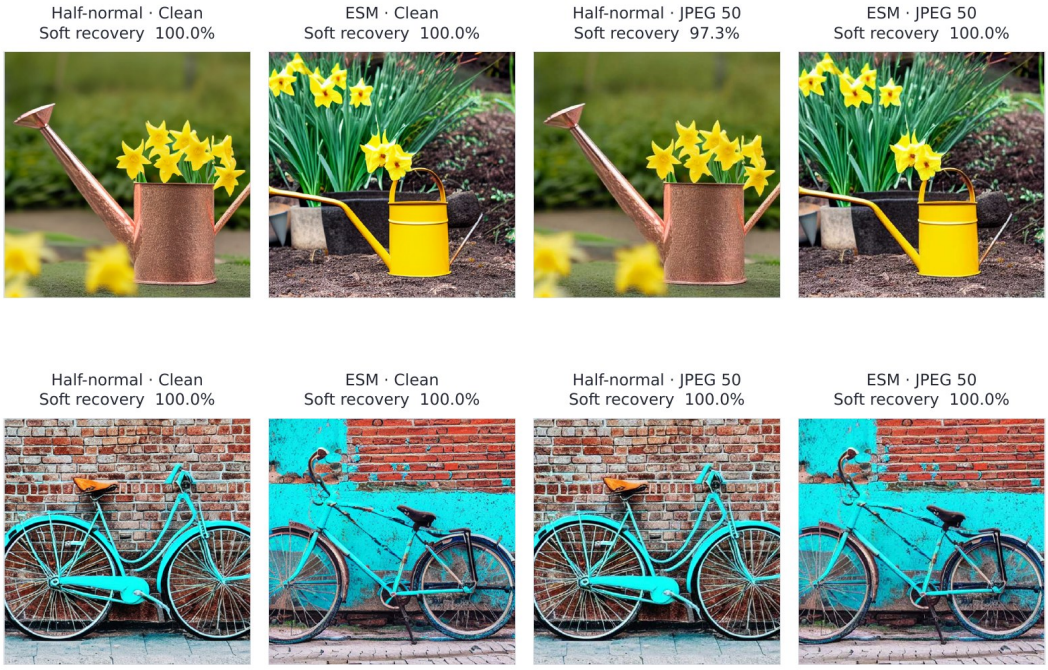


Figure 13: Actual images from the first two fixed cases. Columns pair half-normal and ESM outputs for clean and JPEG50. Prompts and messages match, but images differ; displayed examples do not establish perceptual equivalence. All prespecified gallery cases and channels appear in Appendix D.6.

D.3 Paired design and analysis

Unit of analysis. The primary estimate averages 64 paired image-level accuracy differences with equal weight. Within each key–prompt case, accuracy averages the 1024 correctness indicators; failed generation contributes zero before subtraction. The crossed bootstrap resamples whole key and prompt groups while preserving method pairing and within-image bit dependence. Exact recovery requires every bit to match and is evaluated before rounding accuracy.

Prompt matching, shared payloads and matched block energy control several sources of variation, but the methods still generate different images. This historical experiment therefore measures the combined effect of rotation and magnitude allocation. The separate rotated-HN comparison in the main study isolates amplitude allocation while holding the frame fixed.

Repetition, interleaving and thresholds are fixed throughout this comparison. Its rate counts payload bits, with the nonce supplied externally.

D.4 Connection to the earlier external-state pilot

The earlier external-state pilot used 512 bits with 32 repetitions and dense frames. It returned all 32 images in each primary arm. Under clean/JPEG50, ESM raw accuracy was 93.45/82.56%, versus 82.82/74.24% for HN. Angular-law per-image mean absolute discrepancies were 0.09/0.15 percentage points, using original/recovered latent pairs without a fitted noise parameter. This was a conditional diagnostic on ideal-frame-inspired samples. Post-hoc soft decoding raised JPEG50 exact recovery from 12 to 17 of 32 for HN, and from 26 to 29 for ESM. That observation motivated the historical 1024-bit matched-soft study. Full pilot results, including the deliberately degenerate shared-frame sanity arm and safety-policy amendment, remain in Appendix E.1.

D.5 Complete numerical matrix

All cells contain 64 safe images and zero safety erasures. Delivered and conditional accuracies therefore coincide. Raw accuracy refers to 16384 signs; hard and soft accuracy refer to 1024 payload

bits. Exact-message counts are reported separately because rounding the mean to 100% does not imply complete recovery of every payload.

Table 55: All prespecified nonce-study cells, including the crop failure. HN: half-normal. Exact counts have denominator 64.

		Accuracy (%)			Exact messages	
Channel	**Arm**	Raw	Hard	Soft	Hard	Soft
Clean	HN	83.164	99.689	99.953	19/64	47/64
	ESM	93.707	99.991	99.998	61/64	63/64
JPEG 50	HN	75.358	95.952	97.795	1/64	14/64
	ESM	83.364	98.503	99.254	31/64	47/64
JPEG 30	HN	72.384	93.686	96.126	0/64	7/64
	ESM	79.593	97.263	98.621	19/64	30/64
Resize 256	HN	79.234	98.695	99.519	7/64	26/64
	ESM	89.085	99.886	99.986	48/64	59/64
Crop 410	HN	50.217	50.706	50.763	0/64	0/64
	ESM	50.218	50.690	50.554	0/64	0/64

D.6 Qualitative examples from the 1024-bit study

Cases 0–3 and 16–19 were selected before evaluation. The panels show both methods under all five channels; image labels give soft bit accuracy.

E Earlier external-state pilot

The following study predates the nonce implementation and remains a separate pilot. Its matrix-cache cost is not the current implementation's retained-state cost.

E.1 External-state mechanism pilot

We implemented three matched-rate controls: independent half-normal amplitudes with fresh sign masks; constant-modulus blocks with a reused frame; and constant-modulus blocks with independent ephemeral frames. All carry 512 bits repeated 32 times in a 16384-dimensional latent. Repeats of each bit occupy distinct 128-dimensional blocks. The same message, prompt and underlying normal-amplitude draw are used across methods for each case, preserving block energies up to floating-point error but not identical generated images. These are mechanism controls, not official implementations of Gaussian Shading or Spherical Watermark.

A deliberately restricted sanity control. The shared-frame arm omits the original XOR padding. Its unpermuted repetition leaves at most four distinct normalized block directions in each image, each repeated 32 times. This is a degenerate negative control, not a competitive watermark baseline: its visible collapse and safety-filter frequency must not be attributed to Spherical Watermark. The primary empirical comparison is ephemeral frames versus the half-normal control, both of which return all 32 requested images.

The 32 cases use 16 new authored prompts twice and two shared-frame seeds. SD1.5, fp16, 50-step DPM generation, VAE posterior mode and 50-step DDIM inversion match the earlier pilot. No detector is fitted. This reference implementation uses PCG64 amplitude draws and retains float32 frames as external decoder state; it does not evaluate a robust nonce header.

Safety filtering and sample selection. The safety policy was amended after a flag was observed following two half-normal cases. The same cases and seeds were retained, with the safety checker active and no replacement of flagged images. Flagged requests were excluded from inversion and

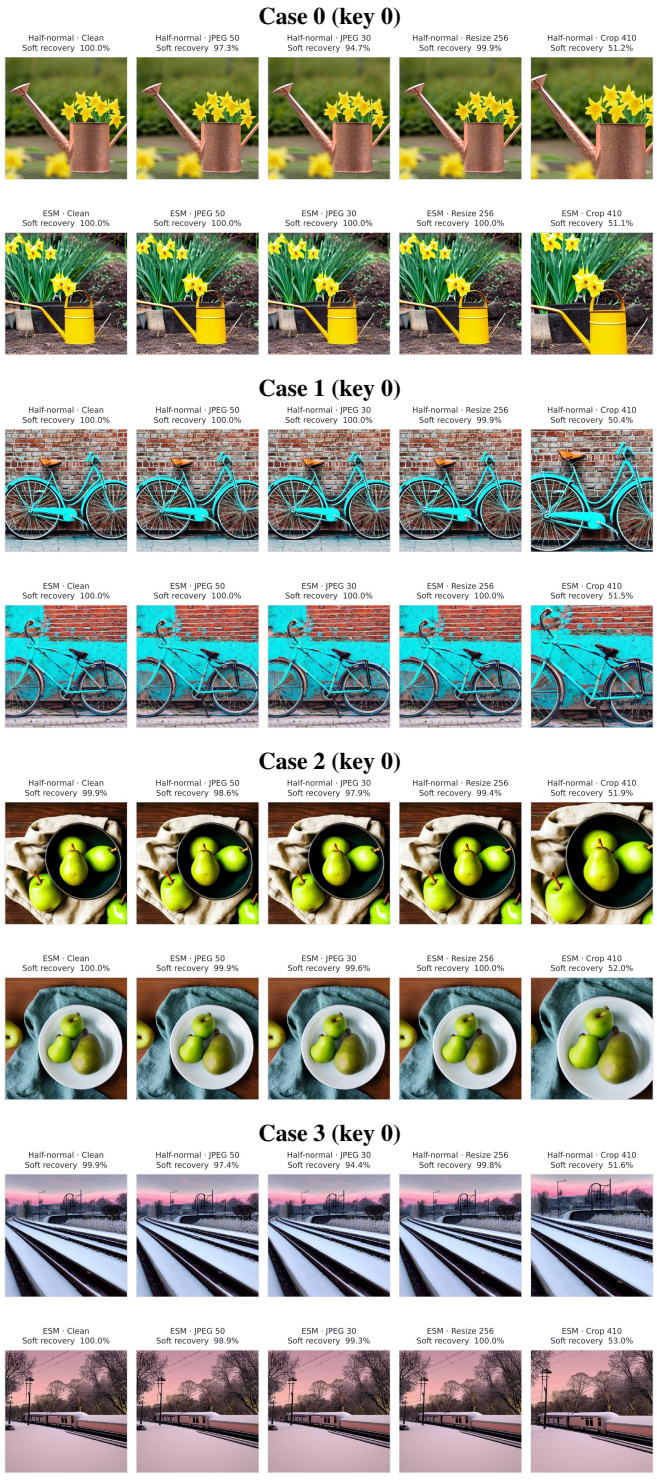


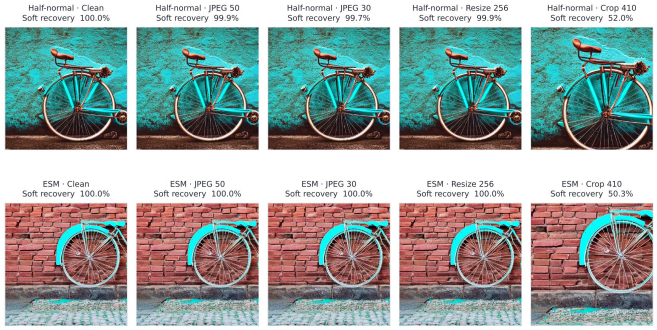
Figure 14: Qualitative examples from the 1024-bit SD1.5 study, continued on the following page. Cases 0–3 (key 0) and 16–19 (key 1) were selected before evaluation. Within each case, the upper row is HN and the lower row is ESM; columns show clean, JPEG50, JPEG30, resize256 and center crop. Each method's generated image supplies all five transformations. Labels report soft bit accuracy.

Figure 14, continued

Case 16 (key 1)



Case 17 (key 1)



Case 18 (key 1)



Case 19 (key 1)



conditional accuracy but counted as zero delivered accuracy. This outcome-informed amendment makes the study exploratory; it is not an independent confirmation under a wholly prespecified selection rule.

Table 56: External-state pilot. Safe is the number of returned safe images out of 32. Raw and payload accuracy are conditional on those safe images. Delivered raw accuracy includes all 32 requests, assigning zero to failures. Exact counts also use all 32. No confidence across a population of secret keys is asserted.

Modulation	Channel	Safe	Raw (%)	Payload (%)	Delivered (%)	Exact
Half-normal	clean	32/32	82.82	99.93	82.82	29/32
	JPEG 50	32/32	74.24	98.46	74.24	12/32
Shared (sanity)	clean	22/32	73.17	98.12	50.30	0/32
	JPEG 50	22/32	68.68	96.90	47.22	0/32
Ephemeral frame	clean	32/32	93.45	100.00	93.45	32/32
	JPEG 50	32/32	82.56	99.35	82.56	26/32

For clean, the ephemeral-minus-half-normal delivered raw-accuracy difference is 10.63 percentage points; a paired 16-prompt-cluster bootstrap gives a 95% percentile interval [9.17, 11.95]. These 2000 bootstrap draws quantify conditional prompt variation in this pilot, not independent key or model replication.

For jpeg50, the ephemeral-minus-half-normal delivered raw-accuracy difference is 8.32 percentage points; a paired 16-prompt-cluster bootstrap gives a 95% percentile interval [5.57, 10.84]. These 2000 bootstrap draws quantify conditional prompt variation in this pilot, not independent key or model replication.

Conditional geometric prediction. We evaluate Theorem 3 using each original/recovered latent pair, with no fitted noise variance. The diagnostic is exploratory and uses the same pilot images.

For clean, observed/predicted mean raw accuracy is 93.45/93.45%; mean absolute per-image discrepancy is 0.09 percentage points. Observed/predicted payload accuracy after interleaved majority decoding is 100.00/100.00%.

For jpeg50, observed/predicted mean raw accuracy is 82.56/82.50%; mean absolute per-image discrepancy is 0.15 percentage points. Observed/predicted payload accuracy after interleaved majority decoding is 99.35/99.34%.

The dense ephemeral-frame implementation takes 210.2 ms mean CPU encoding time per image, including QR, and 2.02 ms decoding given retained matrices. It retains 8.0 MiB of float32 frame state per image. The pilot contains 96 generation requests and 172 inversions; it does not measure frame regeneration or header recovery.

F Historical reusable-frame analysis and detector experiments

This appendix presents the earlier reusable-frame study, including its initial detector failure. It motivates independent block frames and remains separate from the nonce-based recovery experiments.

F.1 Observation models and the reusable frame

Let $k \geq 3$, $C \in O(k)$, $S \in \{-1, 1\}^{k \times B}$, and let $R_b > 0$. A watermarked block is

$$X_b = R_b C S_b / \sqrt{k}, \qquad U_b = X_b / \|X_b\| = C S_b / \sqrt{k}. \tag{7}$$

The signs may be arbitrarily dependent. For the spherical construction, $R_b^2 \sim \chi_k^2$. The implementation studied here uses $k = B = 128$, giving $d = 16384$ coordinates per image. The same C is reused across blocks and images. Its bit encoder repeats 512 message bits 31 times and mixes them with 512 random padding bits before the spherical mapping. Our reproduction follows the sparse mixing structure rather than constructing a dense binary matrix.

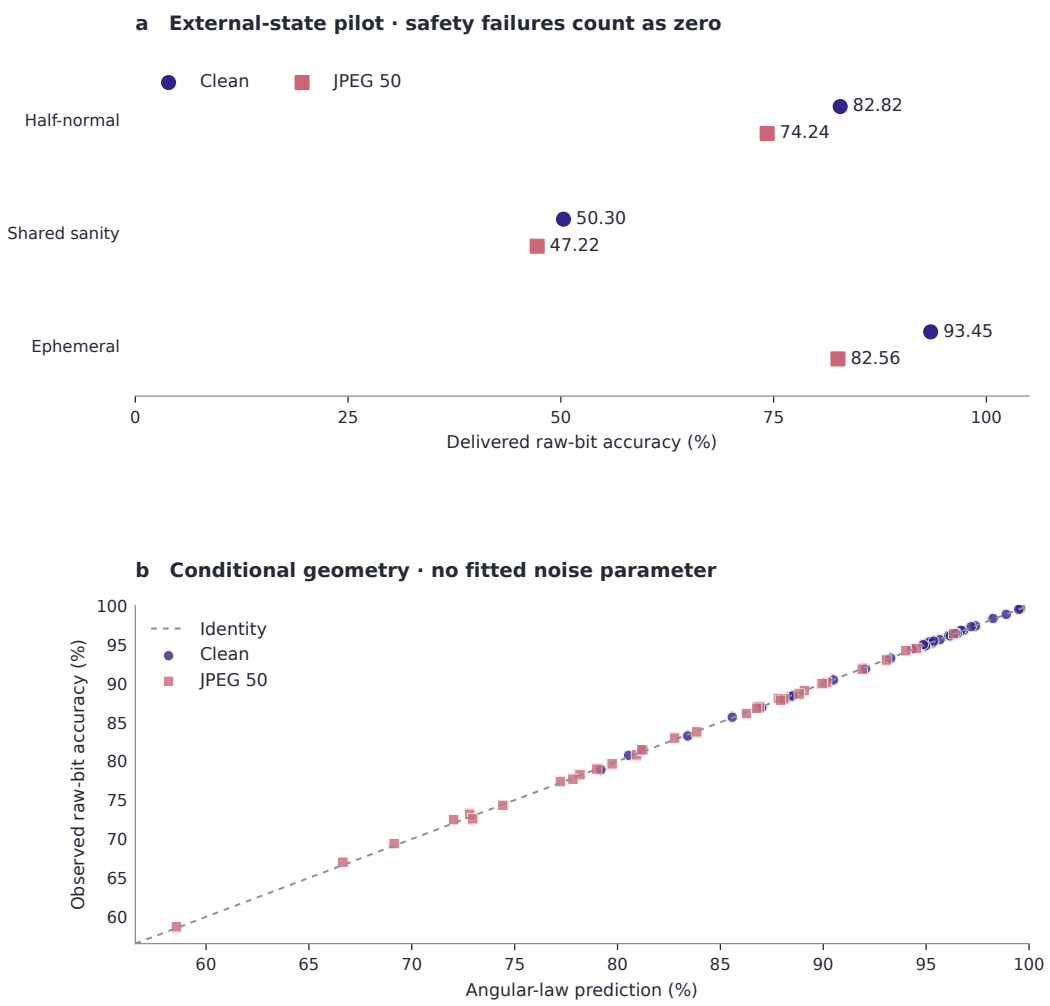


Figure 15: Mechanism and theory diagnostics from the complete external-state pilot. Top: delivered raw accuracy, including generation failures; the repetition-only shared arm is a degenerate sanity control, not Spherical Watermark. Bottom: each safe ephemeral-frame image compared with its parameter-free angular-law expectation. A close fit tests the conditional geometric mechanism; it does not prove finite-key security or image-quality equivalence.

We distinguish three observation models. A *latent observer* sees X directly. A *noisy-latent observer* sees $X + \sigma G$ for independent standard Gaussian G. An *image observer* sees a generated image and applies an available inversion procedure. The last channel is neither assumed Gaussian nor assumed independent of the input. An oracle score may use C only to measure an available signal; a blind score may fit on marked training observations but cannot use C, payloads, or test labels. Known marked training data is a chosen-generator-access assumption, not a mixed-unlabeled-data guarantee.

F.2 Two signatures of a hidden angular alphabet

Theorem 7 (A key-independent Gram lattice)**.** *For any two blocks in* (7)*,*

$$\langle U_a, U_b \rangle = 1 - \frac{2H(S_a, S_b)}{k},$$

where H is Hamming distance. For independent uniform spherical vectors $V, W \in \mathbb{S}^{k-1}$, define

$$T(V, W) = \operatorname{dist}\left(\frac{k(1 - \langle V, W \rangle)}{2}, \mathbb{Z}\right).$$

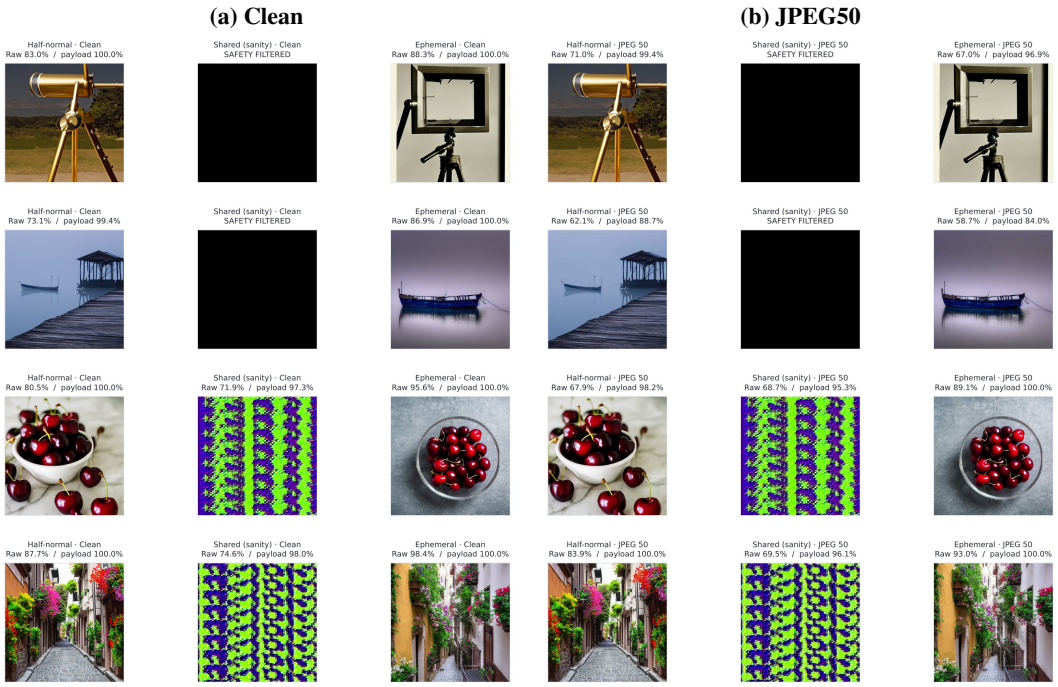


Figure 16: Fixed-index outputs for the first four external-state-pilot cases under (a) clean evaluation and (b) JPEG50. Columns compare the three modulation controls, and labels report raw/payload accuracy conditional on successful generation or after compression. Black panels are actual safety-filtered placeholders, retained as delivery failures rather than replaced; they were not inverted. Prompts and messages match across methods, but the images are not pixel-paired quality controls. The shared-frame arm is the deliberately degenerate sanity control described in the text, not Spherical Watermark.

For $0 \le \tau \le 1/2$,

$$\Pr\{T(V,W) \le \tau\} \le \min\left\{1, \frac{4\tau(k+1)}{k} \frac{\Gamma(k/2)}{\sqrt{\pi}\,\Gamma((k-1)/2)}\right\}.$$

Proof. Orthogonality cancels C from the inner product. Equal signs contribute $1/k$ and unequal signs contribute $-1/k$. Under the null, $Z = \langle V, W\rangle$ has density $f_k(z) = c_k(1-z^2)^{(k-3)/2}$ on $(-1,1)$, where $c_k = \Gamma(k/2)/(\sqrt{\pi}\Gamma((k-1)/2))$. Since $k \ge 3$, this density is bounded by c_k. The event is contained in $k+1$ intervals centered on the lattice, each of length $4\tau/k$. Integrating the density over their union proves the bound. □

The null event at zero tolerance has probability zero. This exact-real statement is not itself a finite-precision cryptographic attack. The tolerance bound is a conservative bridge to measurable implementations, while inversion errors require additional evidence. For perturbed normalized directions $\widehat{U}_a, \widehat{U}_b$, the Gram error is at most $\|\widehat{U}_a - U_a\| + \|\widehat{U}_b - U_b\|$. Consequently the scaled distance can deteriorate proportionally to k. The direct grid signal can therefore be fragile even when payload recovery is excellent.

Theorem 8 (Constant-modulus contrast without independent bits)**.** *For* $W \in O(k)$ *and a spherical random vector* U*, set*

$$J_U(W) = \mathbb{E}\left[k\sum_{j=1}^{k}(W^\top U)_j^4 - 1\right].$$

If $U = CS/\sqrt{k}$ for any distribution on sign vectors, then $J_U(W) \geq 0$ for every W and $J_U(C) = 0$. If U is uniform on $\mathbb{S}^{k-1}$ and independent of W, then

$$J_U(W) = \frac{2(k-1)}{k+2}.$$

Proof. For every unit vector v, Cauchy–Schwarz gives $1 = (\sum_j v_j^2)^2 \leq k \sum_j v_j^4$, with equality precisely when every $|v_j| = 1/\sqrt{k}$. The true frame has this property pointwise. For the spherical null, rotational invariance and $\mathbb{E}U_j^4 = 3/[k(k+2)]$ give $k \sum_j \mathbb{E}U_j^4 - 1 = 3k/(k+2) - 1$. □

This is a population separation, not an identifiability theorem. Different frames may minimize the objective for restricted sign distributions. Nor does the result establish that a local optimization procedure finds a minimizer. It motivates a learned score whose success and sample requirements must be measured on new data. Our pilot uses a whitened FastICA matrix, which is not exactly orthogonal; the theorem motivates that implementation but is not asserted as its finite-sample performance guarantee.

Proposition 9 (Noise decay along a true axis)**.** *Let S_j be a balanced sign, independent of r and G_j, with $r^2 \sim \chi_k^2/k$ and $G_j \sim N(0,1)$. For $Y = rS_j + \sigma G_j$, the standardized excess kurtosis is*

$$\frac{\mathbb{E}Y^4}{(\mathbb{E}Y^2)^2} - 3 = -\frac{2(1-1/k)}{(1+\sigma^2)^2}.$$

Proof. The mean is zero, $\mathbb{E}r^2 = 1$, and $\mathbb{E}r^4 = 1 + 2/k$. Independence and the Gaussian moments give $\mathbb{E}Y^2 = 1 + \sigma^2$ and $\mathbb{E}Y^4 = 1 + 2/k + 6\sigma^2 + 3\sigma^4$. Subtracting $3(1+\sigma^2)^2$ establishes the formula. □

The signal decays smoothly rather than disappearing when a lattice cell is crossed. However, an observer still has to learn a useful direction. An oracle signal calculation cannot replace that learning experiment.

F.3 What would a distributional repair require?

Randomizing one common radius does not fill the angular space. More generally, a repair must account for both the sign law and the magnitude law. Let $\gamma_d = N(0, I_d)$, let ν_d be the product of d half-normal distributions, and measure entropy in bits.

Theorem 10 (Exact sign–magnitude decomposition)**.** *Fix $C \in O(d)$. Let $S \in \{-1,1\}^d$ and $A \in (0,\infty)^d$ have any joint law, and put $Y = C(S \odot A)$. Then, with extended-real relative entropy,*

$$D_{\mathrm{KL}}(P_Y \| \gamma_d) = D_{\mathrm{KL}}(P_S \| \operatorname{Unif}\{-1,1\}^d) + \mathbb{E}_S D_{\mathrm{KL}}(P_{A|S} \| \nu_d).$$

If A and S are independent, this is

$$D_{\mathrm{KL}}(P_Y \| \gamma_d) = (d - H(S)) \log 2 + D_{\mathrm{KL}}(P_A \| \nu_d). \tag{8}$$

Proof. The map $y \mapsto (\operatorname{sign}(C^\top y), |C^\top y|)$ is a measurable bijection outside the rotated hyperplanes $\{y : (C^\top y)_j = 0\}$, which have probability zero under both laws. Relative entropy is invariant under this bijection. Orthogonal invariance of γ_d transforms its law into the product of uniform signs and independent half-normal magnitudes. The chain rule for relative entropy proves the first identity. Independence gives the second, using $D_{\mathrm{KL}}(P_S \| \operatorname{Unif}) = d \log 2 - H(S) \log 2$. □

Consequences. Independent half-normal magnitudes remove the magnitude term exactly, but cannot remove a sign entropy deficit. An injective binary encoder of h uniform input bits has $H(S) = h$. For the reproduced random-message encoder, $h = 512 + 512 = 1024$, so even a perfect independent-magnitude completion leaves $15360 \log 2$ nats of conditional-on-key relative entropy. For a fixed message, only the random padding contributes entropy. The original common-radius model is singular: within each block all pre-rotation absolute values are equal, an event of zero Gaussian measure. Its relative entropy is therefore infinite in the ideal real-valued model.

These facts concern a fixed key and an information-theoretic divergence. They do not determine computational detectability or perceptual image quality. Data processing only propagates closeness forward; it does not force every latent distinction to survive a lossy generator.

F.4 The padding encoder is a parity graph

There is a more precise reason not to replace three-wise independence by independent coordinates. Make a bipartite graph with message vertices on one side, padding vertices on the other, and an additional fixed-sign anchor on the message side. Every mixed bit is an edge between its message and padding vertex; every appended padding bit is an edge from that padding vertex to the anchor. Up to a fixed sign convention, encoded signs are products of independent vertex signs.

Theorem 11 (Cycle leakage and fourth moments)**.** *Let $G = (V, E)$ be a finite simple bipartite graph. Assign independent uniform signs B_v to its vertices, possibly fixing one anchor sign, and set $S_{\{u,v\}} = B_u B_v$. Then every subset of at most three edge signs is jointly uniform. For any edge set F,*

$$\mathbb{E} \prod_{e \in F} S_e = \mathbf{1}\{\textit{every vertex has even degree in } (V, F)\}.$$

If $c(G)$ is the number of connected components, $H(S) = |V| - c(G)$ bits. Writing $\mathcal{C}_4$ for the sets of four edges forming a cycle, any $a \in \mathbb{R}^{|E|}$ satisfies

$$\mathbb{E}(a^\top S)^4 = 3\|a\|_2^4 - 2\|a\|_4^4 + 24 \sum_{F \in \mathcal{C}_4} \prod_{e \in F} a_e.$$

In particular, independent half-normal completion does not remove cycle leakage: for $Y_e = S_e A_e$ and a four-cycle F, $\mathbb{E} \prod_{e \in F} Y_e = (2/\pi)^2$, despite each Y_e being marginally standard Gaussian and each pair being independent.

Proof. In the edge product, each vertex sign appears with exponent equal to its degree. Averaging kills the product if any unfixed vertex has odd degree. If all unfixed degrees are even, a possibly fixed anchor must also have even degree by the handshaking identity. A nonempty even-degree subgraph is a union of cycles. A simple bipartite graph has no cycle of length below four, so all nontrivial joint sign moments for up to three distinct edges vanish, which characterizes their uniform joint law.

Choose a spanning forest. Its $|V| - c(G)$ edge signs can be assigned arbitrarily and are uniform, since each assignment has an equal number of vertex-sign preimages. Every other edge sign is the product of forest signs along its path. This proves the entropy formula. Expanding the fourth power leaves repeated-index terms $\sum_e a_e^4 + 6 \sum_{e<f} a_e^2 a_f^2$ and distinct four-edge cycles, each with coefficient $4! = 24$. These are exactly the displayed terms. Finally, $\mathbb{E}A_e = \sqrt{2/\pi}$, and independent magnitudes factor across a cycle. Balanced single signs and independent sign pairs establish the marginal and pairwise claims. □

Exact integer calculations for keys 0, 1, and 2 give total encoder rank 1024. Their 128-edge blocks have ranks ranging from 112 to 127, 112 to 126, and 112 to 127 respectively. Mean numbers of four-cycles per block are 10.6016, 10.8281, and 10.3359. These are computations for the inspected encoder instances, not population averages over keys. The theorem identifies the fourth-order dependence missed by an independent-bit ICA model, and shows why merely replacing the radii by independent magnitudes is not a complete repair.

F.5 Learning and held-out calibration

Training, calibration, and testing are separate operations. Given marked training latents or inversions, we flatten the 128-dimensional blocks and fit symmetric cubic FastICA after empirical whitening. We fit for at most 150 iterations, with float64 SVD-based orthogonalization. On a new observation, we normalize each block and compute negative mean fourth power in the learned coordinates. Higher scores indicate the constant-modulus structure. Neither the true frame nor the payload participates in fitting or testing. We report frame alignment only as a post-hoc diagnostic in the synthetic setting.

Proposition 12 (Finite-sample null calibration)**.** *Conditional on the training data, let $s_1, \dots, s_n, s_*$ be exchangeable null scores from a fixed learned detector. Then*

$$p_* = \frac{1 + \sum_{i=1}^{n} \mathbf{1}\{s_i \geq s_*\}}{n+1}$$

satisfies $\Pr(p_* \leq \alpha) \leq \alpha$ *for all* $\alpha \in [0, 1]$.

Proof. Break ties uniformly at random. By exchangeability, the descending rank of s_* is uniform on $\{1, \dots, n+1\}$. Counting every tie against rejection gives a p-value no smaller than its randomized-rank counterpart. Its lower tail is therefore bounded by $\lfloor \alpha(n+1) \rfloor / (n+1) \leq \alpha$. □

The requirement is exchangeability after fixing the detector. It fails if null test data influence training or if the target domain shifts relative to calibration. Our small feasibility pilot reports AUC rather than pretending to estimate 10^{-3} or 10^{-6} false-positive rates.

Monotone optimization after the failed pilot. The revised detector normalizes training blocks, centers them, and whitens their empirical covariance. For the resulting row matrix Z, it minimizes

$$F(W) = \frac{1}{4N} \sum_{i,j} (ZW)_{ij}^4, \qquad W \in O(k).$$

Writing $G = Z^\top (ZW)^{\odot 3}/N$, we project the gradient onto the tangent space as $G - W \operatorname{sym}(W^\top G)$ and retract by the polar factor. Up to 15 backtracking trials seek Armijo decrease; a candidate that increases the objective is rejected. The iteration stops at 400 steps, a rejected step, or a squared tangent-gradient norm below 10^{-9}. This is a standard Riemannian optimization construction, related to established fourth-power dictionary-learning objectives Zhai et al. (2020); its use here is a detector implementation, not a new general-purpose optimizer. The population theorem does not guarantee recovery under arbitrary image inversion.

F.6 Synthetic protocol and results

The first pilot uses three independently sampled keys, with 128 marked and 128 Gaussian samples per key. Independent Gaussian noise of the same standard deviation is added to both arms. At zero noise, every payload bit is recovered. Table 57 reports arithmetic mean AUC across keys; this exploratory mean is not a confidence interval.

Table 57: Exploratory latent detection. ICA is trained on 256 different marked images per key, using training noise 0.5; its test set has 128 images per arm per key. The grid uses a separate pilot. No row is an image-space result.

Test noise σ	Gram-grid AUC	Learned ICA AUC
0.00	1.0000	1.0000
0.03	1.0000	not run
0.10	0.4853	1.0000
0.30	0.5043	1.0000
0.50	0.5308	1.0000
1.00	0.5028	0.9771
1.50	not run	0.9156

At training noise 0.5, mean best-axis alignments for keys 0, 1, and 2 are respectively 0.9989, 0.3608, and 0.9989. This heterogeneity is essential: strong mean detection does not establish uniform frame recovery. With clean training, all three alignments are approximately 0.9995. The direct lattice test becomes ineffective near $\sigma = 0.1$, while the learned fourth-order score retains signal at larger noise levels. Payload bit accuracy remains high in this synthetic channel, which demonstrates that payload robustness and statistical hiding are different axes.

An initial float32 eigendecomposition-based variant produced nonfinite iterates in two noisy-training runs. Those runs are excluded from the detection results. The reported ICA variant uses float64 SVD retraction and rejects nonfinite iterates.

F.7 IMAGE-SPACE FEASIBILITY PROTOCOL

The image protocol uses SD 1.5, 128 marked training images and 32 held-out images per test arm, one fixed key, and 16 authored benign prompts. Generation uses 50 DPM-Solver multistep steps with guidance 7.5; inversion uses 50 DDIM inverse steps with empty prompts and guidance 1. Images are quantized to PNG and encoded with the VAE posterior mode. The pilot uses fp16 and posterior-mode encoding, whereas the source implementation uses fp32 and posterior sampling. The primary metric is held-out learned-score AUC, accompanied by grid AUC, inversion RMSE and payload accuracy.

Initial detector failure. The fixed image-space detector obtains AUC 0.4805; the Gram-grid detector obtains 0.4834 on 32 held-out marked and 32 unmarked images. Neither predefined detector distinguishes the two arms in this pilot. This failure does not establish watermark security: it motivates the exploratory diagnostics below, which reuse the observed test set.

Post-hoc localization. The true-frame oracle obtains AUC 1.0000 on the same image inversions, while the failed learned matrix obtains AUC 0.4170 even on their original, pre-generation latents. This points to a failure of frame learning rather than loss of all signal during inversion. Retraining on normalized blocks with three new initializations yields AUC 0.6006, 0.6270, and 0.8350. These results remain exploratory because the test set was already observed. Mean payload bit accuracy on the 32 marked test images is 99.8291%, and 28/32 payloads are recovered completely.

The monotone quartic optimizer subsequently obtains AUC 1.0000 from each of three initializations on this same pilot. A second protocol freezes a single initialization and all optimizer constants, then uses new keys 11 and 12, new latent RNG seeds, and 16 new prompts. Each key again has 128 marked training images and 32 test images per arm. The detector obtains AUC 1.0000 on each new key; its fits stop after 144 and 129 iterations, respectively. These are two small independent confirmations under the declared protocol, not a confidence interval across the population of keys. Prompts are new relative to the exploratory pilot but repeat within each run. Across both stages, 576 images are generated and inverted. The old-set optimization result remains exploratory and is not pooled with the fresh tests.

Image-channel diagnostic with a frozen detector. The channel diagnostic evaluates 640 inversions and eight round-trip generations using the same 128 held-out originals and two fitted detectors. Reusing these images makes this a conditional robustness analysis rather than an independent confirmation. Under JPEG quality 50, AUC is 0.9961/0.9912 and marked-image bit accuracy is 96.31%/94.19% for keys 11/12, respectively. Under downsampling to 256 pixels and back, AUC is 1.0000/1.0000 and marked-image bit accuracy is 99.60%/98.67% for keys 11/12, respectively. Under a centered 410-pixel crop and resize, AUC is 0.4766/0.4990 and marked-image bit accuracy is 50.20%/49.82% for keys 11/12, respectively. Table 58 reports all conditions; the appendix displays fixed-index channel outputs, round trips, and an explicitly outcome-selected failure gallery. These measurements concern the reproduced construction and this inversion pipeline, not all diffusion watermarks or all image transformations.

F.8 EXTERNAL-STATE MODULATION PILOT

A separate 96-request pilot compares ephemeral frames with matched-rate modulation controls (Appendix E.1). Conditional raw-bit accuracy is 93.45/82.56% for ephemeral frames under clean/JPEG50, versus 82.82/74.24% for the half-normal control. The results distinguish conditional accuracy, delivered accuracy and repetition decoding. The angular law in Theorem 3 relates the observed latent geometry to decoding errors without assuming AWGN. This pilot uses external frame state and does not evaluate nonce recovery or complete official baseline systems.

G HISTORICAL IMAGE EXAMPLES AND CHANNEL DIAGNOSTICS

The panels show generated images and their specified transformations from the earlier detector study. Figure 17 and the gallery below show the first eight key-11 test prompt pairs in order. These examples were selected by index; the separate failure gallery explicitly uses outcome-based selection.

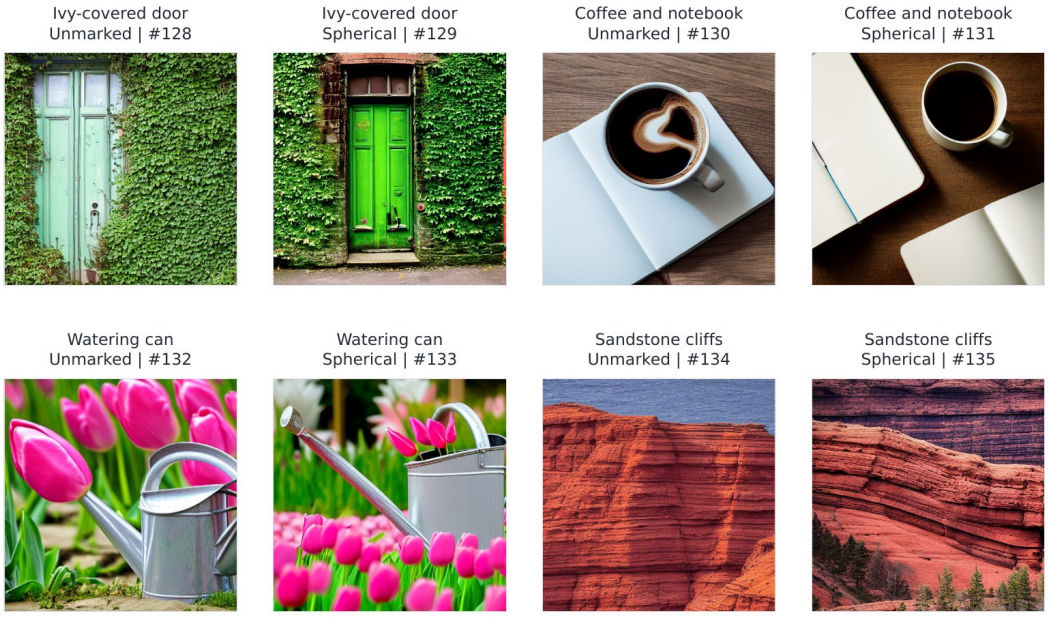


Figure 17: SD1.5 outputs for the first four test prompt pairs of key 11 (indices 128–135), selected by index. Each pair uses the same prompt but different latent draws: unmarked on the left and Spherical Watermark on the right. These examples do not establish paired perceptual equivalence. Additional samples and channel diagnostics appear in Appendix G.

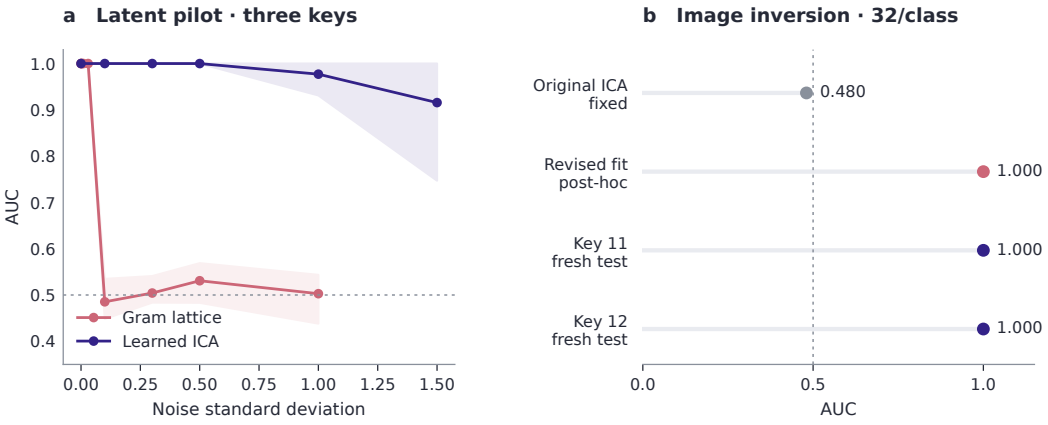


Figure 18: Evidence levels are distinct. (a) Synthetic means with minimum–maximum bands across three keys. (b) An initial image-space failure, post-hoc optimization on the old data, and fresh-key confirmation are separate experiments; test sample size is 32 per arm per key. The separate external-state candidate pilot is reported in Appendix E.1, not pooled into these detector results.

Image-channel diagnostic. We evaluate all 128 held-out images from keys 11 and 12 (32 marked and 32 unmarked per key). Each image receives five deterministic channels: clean; JPEG qualities 90 and 50 with 4:2:0 chroma subsampling; Lanczos downsampling to 256×256 and back; and a centered 410×410 crop followed by Lanczos resizing to 512×512. The crop retains approximately 80% of each side, or 64% of the area. Both classes receive identical channels. Each key's detector remains fixed across the 640 observations, without channel-specific refitting or threshold selection.

The diagnostic applies the clean experiment's SD1.5 model and inversion settings to the same previously inspected test images, allowing comparison across channels for fixed images and keys.

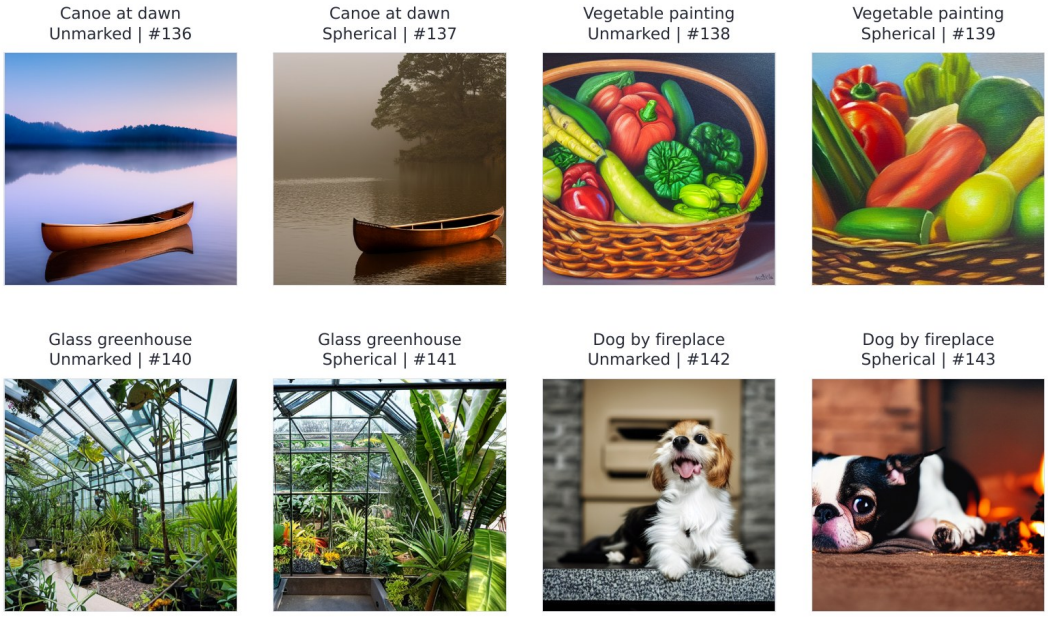


Figure 19: The next four test prompt pairs for key 11 (indices 136–143). Unmarked and marked outputs alternate; each pair shares a prompt and uses independent latent draws.

Channel results. The comparison contains 640 image inversions and eight additional round-trip generations. Table 58 reports detection and payload recovery separately for each key. With 32 test images per class and two keys, these measurements do not resolve rare false positives or population-level variation across keys.

Table 58: Image-channel diagnostic. BA is mean marked-image bit accuracy; exact counts are complete 512-bit payloads out of 32. Both classes receive the same transform, and the learned detector is held fixed.

	Key 11			Key 12		
Channel	AUC	BA (%)	Exact	AUC	BA (%)	Exact
Clean	1.0000	99.98	30/32	1.0000	99.57	28/32
JPEG 90	1.0000	99.54	26/32	1.0000	98.55	24/32
JPEG 50	0.9961	96.31	13/32	0.9912	94.19	10/32
Resize 256	1.0000	99.60	23/32	1.0000	98.67	20/32
Crop 410 pixels	0.4766	50.20	0/32	0.4990	49.82	0/32

G.1 A SECOND KEY AND THE ORIGINAL NEGATIVE PILOT

H ADDITIONAL RELATED WORK

The main text introduces the mechanisms closest to ESM. This appendix places them in the broader literature by following three design choices: where a watermark enters the image pipeline, what information its decoder receives, and which recovery or security property it targets.

H.1 LEARNED IMAGE-SPACE AND MODEL-ROOTED WATERMARKS

HiDDeN learns an image encoder and decoder through a distortion layer, recovering embedded data after transformations such as compression and cropping (Zhu et al., 2018). StegaStamp trains against perturbations that approximate physical image formation, extending learned embedding toward print–capture recovery (Tancik et al., 2020). Both allocate redundancy through training in image

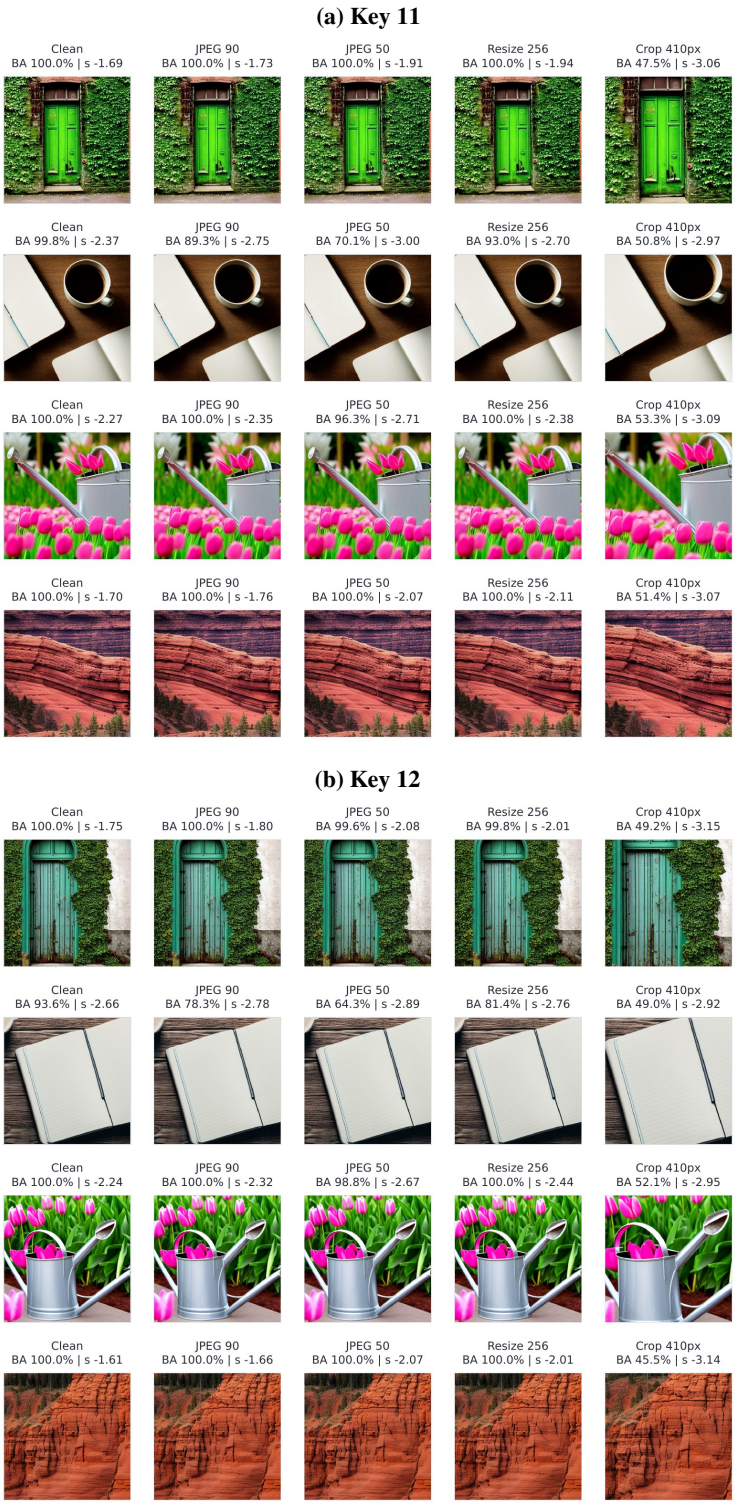


Figure 20: Actual channel outputs for (a) key 11 and (b) key 12, using marked indices 129, 131, 133, and 135 selected before channel outcomes. Columns show clean, JPEG 90, JPEG 50, resize, and crop. BA is authorized payload accuracy; s is the frozen blind detector score (larger means more marked-like). A visually mild change can affect decoding and statistical detection differently. These fixed-index images illustrate, but do not replace, the full table.

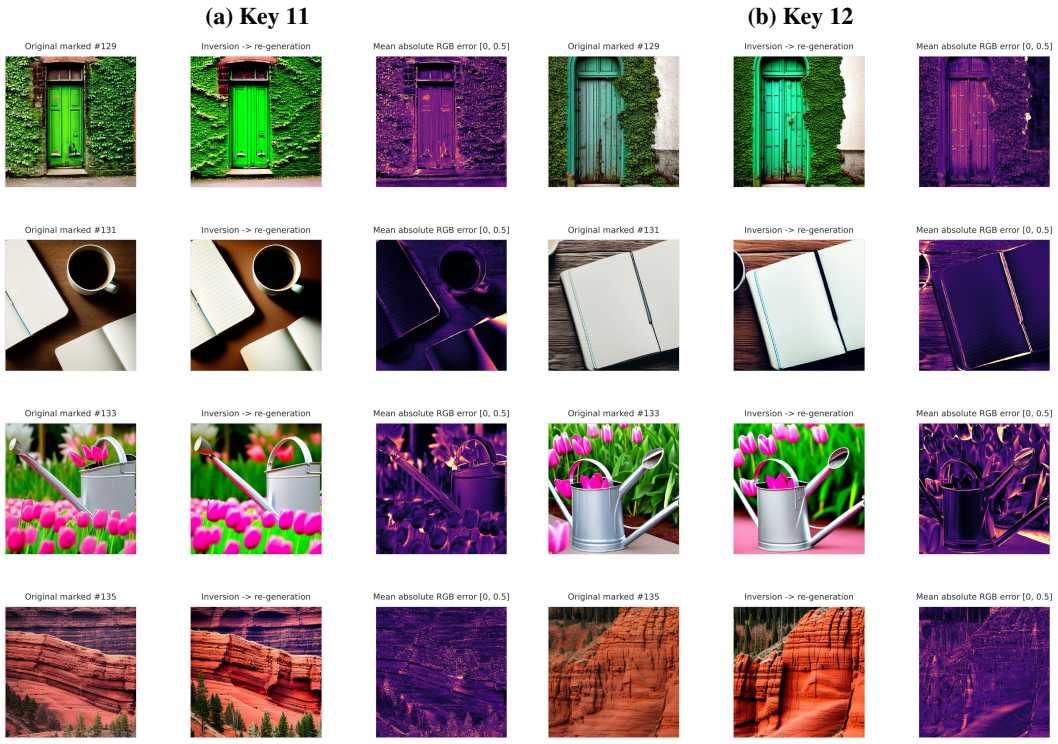


Figure 21: Round-trip diagnostics for (a) key 11 and (b) key 12: the same four fixed-index marked originals, re-generation from their clean empty-prompt DDIM inversions using the original prompts, and mean absolute RGB difference maps. Error is divided by 255 and shown on one fixed scale from 0 (black) to 0.5 (pale yellow), saturated above 0.5. The maps are diagnostic transformations, not photographs. Round-trip images use 50 DPM-Solver steps and guidance 7.5. This is not an optimized removal attack or an image-quality comparison between watermark methods.

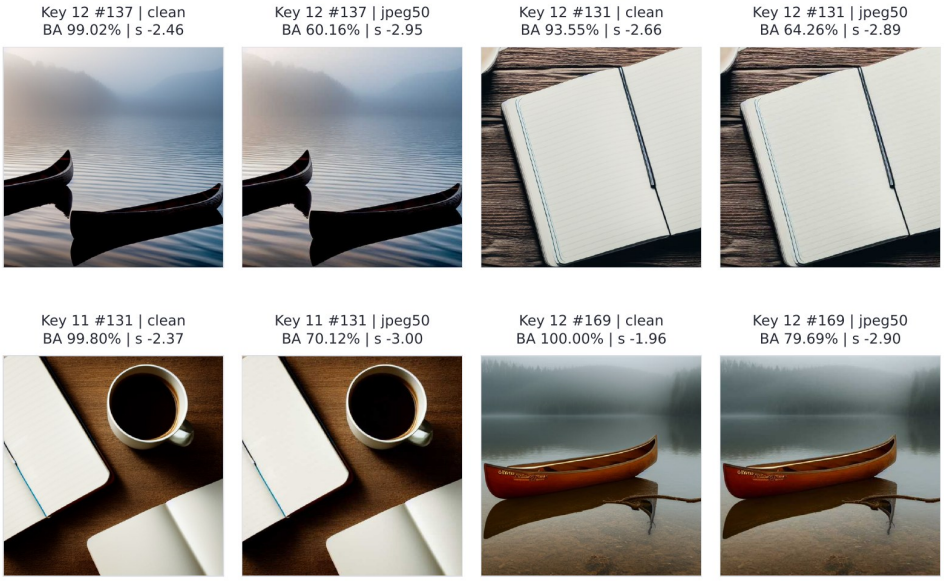


Figure 22: Outcome-selected failure inspection: the four lowest JPEG-50 payload accuracies across both keys, with clean and compressed images alternating. Ties are broken by key then index. Unlike the fixed-index galleries, this panel is deliberately not representative. All 64 marked JPEG-50 cases remain in the numerical table. Poor authorized decoding does not by itself imply failure of blind detection, or vice versa.

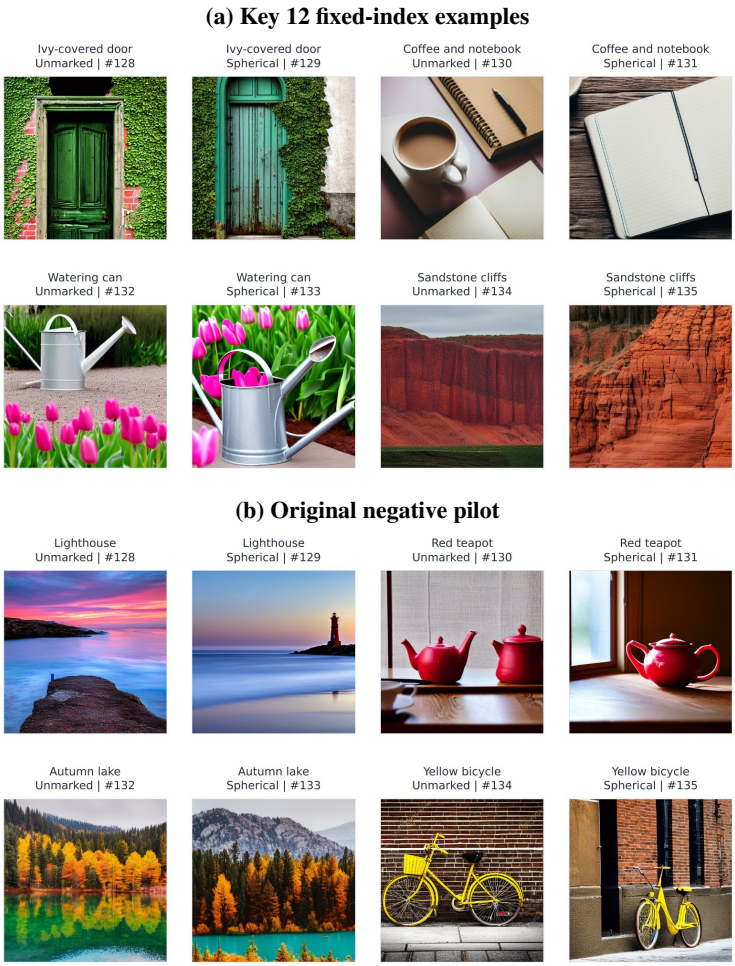


Figure 23: Historical image examples with alternating unmarked and marked outputs. (a) Key 12, first four test prompt pairs (indices 128–135), added after inspection of the clean experiment. (b) Key 0, first four test prompt pairs from the original negative pilot, whose predefined learned detector had full-test AUC 0.4805. Each prompt pair uses independent latent draws.

space. ESM specifies the latent amplitudes and coordinate frames directly, allowing embedding in an existing diffusion generator without training a watermark network.

Stable Signature embeds a signature by fine-tuning the latent decoder of a diffusion model (Fernandez et al., 2023). Watermark Anything learns localized embedding and extraction, including identification of watermarked regions and recovery of multiple messages from a composite image (Sander et al., 2025). These methods illustrate how the target shapes the decoder: a global model signature, a repeated payload and local provenance call for different information and decision rules.

H.2 Initial-noise structure and distribution-preserving modulation

Tree-Ring embeds a pattern in the Fourier domain of the initial noise and tests for it after diffusion inversion (Wen et al., 2023). RingID develops a heterogeneous multi-channel construction for multi-key identification (Ci et al., 2024). Their emphasis on distinguishing candidate keys complements the payload-recovery setting, where the decoder receives the appropriate key and reconstructs the message.

Gaussian Shading targets distribution-preserving latent embedding (Yang et al., 2024a), while Gaussian Shading++ develops nonce handling for practical deployment (Yang et al., 2025). Spherical Watermark combines a binary angular construction with random radii (Hu et al., 2026); correlation-aware angular watermarking further examines dependence in angular codes (Zheng et al., 2026).

ESM focuses on the frame assignment and its decoding geometry. The common-pipeline experiments use source-based GS and SW adapters, with details in Appendix A. The HN and rotated-HN controls clarify the roles of frame rotation and coordinate magnitudes.

H.3 Repeated observations and frame structure

Repeated observations expose structure that may be weak in a single sample. Yang et al. (2024b) study averaging-based steganalysis of content-agnostic watermark patterns, including removal and forgery in their specified settings. Our reusable-frame analysis follows the same broad concern through Gram invariants and fourth-order contrasts. Independent-component analysis and orthogonal fourth-order optimization supply tools for learning such structure (Hyvärinen, 1999; Zhai et al., 2020), while classical Haar sampling supplies the ideal frame construction (Mezzadri, 2007).

H.4 Generation, inversion and channel dependence

Latent diffusion separates image compression from the generative diffusion process (Rombach et al., 2022). DDIM provides a deterministic diffusion formulation used for reconstruction and inversion (Song et al., 2021); DPM-Solver and DPM-Solver++ develop efficient diffusion ODE solvers, including guided sampling (Lu et al., 2022a;b). These choices enter watermark recovery through the latent estimated from the final image. Changing the sampler, guidance, inversion procedure or autoencoder can change the recovered block angles.

The conditional angular law provides a way to examine those changes. Given a source/recovered pair, it translates the longitudinal and transverse components into a distribution of decoding outcomes. The image pipeline determines the pair; the random-frame geometry determines the conditional law. This separation lets the same geometric analysis be compared with several measured channels.

H.5 Robustness benchmarks and statistical units

WAVES evaluates detection and identification under diverse distortions, diffusion-based changes and adversarial stress tests, alongside image-quality degradation (An et al., 2024). It motivates joint reporting of watermark survival and the cost of an attack to image content. The five-channel matrix in our main study provides a matched comparison of compression, resizing and cropping, with complete-message recovery as the image-level endpoint.

For uncertainty estimates, the statistical units follow the way images are generated. The same keys recur across prompts, and the same prompts recur across keys. Owen's pigeonhole bootstrap resamples row and column factors separately for such crossed data (Owen, 2007). We therefore resample key and prompt groups while retaining method pairing. Bit accuracy summarizes each image's coordinate decisions, and exact-message recovery records whether the entire payload is recovered.

I On the epigraph

The epigraph on the first page is from Poincaré's *Science and Hypothesis* (Poincaré, 1905, Chapter 2): "Mathematicians do not study objects, but the relations between objects; to them it is a matter of indifference if these objects are replaced by others, provided that the relations do not change." We use it as a one-sentence summary of the paper's viewpoint, not as an argument; the statements it summarizes are Propositions 1 and 2 and Theorem 3.

The objects and the relations. A watermark's objects are its codeword x and its secret frames C_P. Under (A1) the diffusion pipeline is in the position of Poincaré's mathematician: it receives the latent X, and by Proposition 1 the law of X depends on the code only through the relations among the blocks that share a frame, the Gram invariant $\Gamma_P(x) = (x_b^\top x_c)_{b,c\in P}$. Replacing a codeword by another with the same Gram invariant, or a frame by another in its fiber, changes nothing the pipeline can observe. With one frame per block the only relations left are the block norms, so every prior-preserving code hands the pipeline the same Gaussian and, by Proposition 2(i), the same joint law of transmitted and recovered latents.

What the decoder reads is a relation. The decoder also sees each block only through a relation. Conditional on the latent, the frame is uniform on its fiber, and the decoded block depends on the recovered block Y_b only through its component a_b along X_b and the length h_b of the remainder (Proposition 2(iii)). For a given magnitude profile, the error law of Theorem 3 therefore depends on the image, the model and the transformation only through the recovered angle ψ_b between a source block and its recovery. This is what licenses the channel transplant of Section 5.1: pairs (X, Y) measured with one watermark's images can replace another's without changing any relation the decoder uses, and there they reproduce the observed recovery rates without fitted parameters, with the two exceptions reported in that section.

When the relations change. The proviso of the quotation marks where each statement stops. If blocks share a frame, their inner products are a relation the pipeline can see, and equal magnitudes place them on a lattice (Proposition 1(ii)); this is why Spherical Watermark's key-fixed frame and a reused ESM nonce leave the latent singular with respect to $N(0, I)$, while ESM with fresh nonces leaves it exactly Gaussian. If the channel destroys the relation between a source block and its recovery, as the tested crop does by leaving recovered blocks nearly orthogonal to their sources (Section 5.3), every readout of the family is at chance, whatever the code. If the channel could observe the frames themselves, contrary to (A2), the objects would matter again and the fiber argument would no longer apply.

A classical form. In statistics the same idea is the principle that a law invariant under a group is determined by the law of a maximal invariant (Eaton, 1989, Section 4.1). For a frame shared by the blocks of a part P, two tuples of blocks lie in the same orbit exactly when they have the same Gram matrix, so Γ_P is the maximal invariant, and Proposition 1 is this principle applied to the groups generated by frame sharing.