RECEIVER-DISTRIBUTION-BASED ALPHABET DESIGN FOR STRUCTURED-LIGHT COMMUNICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Due to the deteriorating effect of atmospheric/oceanic turbulence on optical signals, the design of laser source profiles plays an important role in free-space optical communication. Alphabet design becomes difficult in such random media when separation between transmitted fields does not reliably translate into separable receiver distributions and no tractable receiver-side criterion is available. We introduce receiver-distribution alphabet design, which learns operational pairwise distinguishability from held-out cross-entropy induced by classifiers with calibration on receiver-intensity observations. The resulting empirical scores support finite-pool selection under either of two objectives defined on the same scores: a max–min rule that protects the least distinguishable selected pair, and a min-sum rule that minimizes the total pairwise loss, without assuming an analytic channel metric. We evaluate this principle in physics-based simulations of structured-light propagation through random media, and we ask which of the two objectives decodes better under this score. On propagation realizations disjoint from selection, min-sum wins: it improves mean classification accuracy by 13.80 percentage points over an unadapted binary reference, and by 4.39 points over max–min. These results support receiver-distribution-aware alphabet design when source-space geometry alone does not reliably predict receiver separability.

1 INTRODUCTION

Alphabet design asks how to choose a finite transmit alphabet whose received observations remain distinguishable. Classical design approaches optimize an analytic criterion that a tractable channel model ties to error probability: a divergence between class-conditional receiver distributions (Kailath, 1967), a pairwise-error bound determined by source-space geometry (Hochwald & Marzetta, 2000), or a packing distance between source signals (Love et al., 2003). However, in structured-light communications (Avramov-Zamurovic et al., 2023), the transmitter controls the source wave fields while the receiver observes distorted intensity images. The distortion is described by a paraxial model of propagation through a random medium, and intensity-only detection means that the source-space separation does not imply separable receiver distributions.

Prior work on structured light and orbital-angular-momentum (OAM) communication has studied transmission, classification, turbulence correction, and generative augmentation for prescribed or heuristically chosen source mode sets (Krenn et al., 2014; Doster & Watnik, 2017; Lohani & Glasser, 2018; Wang et al., 2026). Work that designs the alphabet itself has emphasized diversity among source intensity profiles (Jarrett et al., 2024). Coefficient separation and source-intensity overlap are distinct quantities, but criteria of either kind are defined at the transmitter and cannot capture how random propagation and intensity detection reshape class separability at the receiver. Cisternas & Anguita (2024) use a prescribed optimal-transport distance between modeled receiver spectra to select OAM constellations. This addresses the receiver-side structure, but its selection criterion is tied to a specified spectral representation and optimal-transport metric. Their score summarizes each class by a barycenter spectrum and a mean dispersion around it, so it separates classes only as far as those two summaries differ. It also needs a phase-sensitive observable: its transport acts along the topological-charge line, and forming an OAM spectrum requires transverse phase that intensity images do not retain. End-to-end systems jointly learn a constellation and receiver through a differentiable channel model (O'Shea & Hoydis, 2017; Stark et al., 2019). We do not assume one. The criterion then applies to a channel available only as a black-box simulator or as measurement. It also

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.

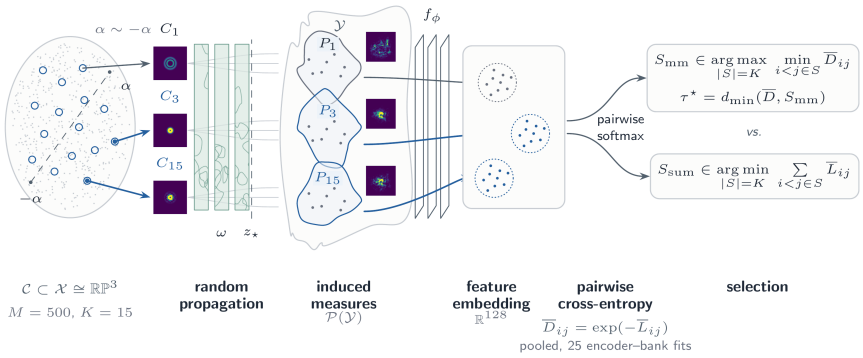

Figure 1: Overview of receiver-distribution alphabet design. A d-mode coefficient space is discretized into M candidates; simulated receiver observations define classifier-induced pairwise scores used for finite-pool selection of K symbols, which are evaluated on propagation realizations disjoint from selection. Here, $\mathcal{C} \subset \mathcal{X}_d$ is the candidate pool (with $\alpha \sim -\alpha$ identifying intensity-equivalent coefficients), ω indexes a random-medium realization, P_i is the induced receiver-intensity distribution for candidate i, and f_ϕ is the learned feature encoder. The matrices $\overline{L}$ and $\overline{D}$ are the pooled pairwise cross-entropy loss and distinguishability, respectively; S_{mm} and $\tau^\star$ denote the exact max–min alphabet and its optimal bottleneck value, and S_{sum} denotes the min-sum alphabet.

avoids backpropagating through a long sequence of random phase screens, which is memory-bound and gradient-noisy. Thus, these approaches do not directly provide a design criterion for selecting, from a physically specified finite family, symbols whose full receiver-intensity distributions are empirically distinguishable in random media. We instead perform receiver-distribution alphabet design over a physically specified candidate family, requiring no tractable pairwise error criterion.

For superpositions of d basis modes, the intensity-equivalent coefficient space is $\mathbb{RP}^{d-1}$. We discretize this space into M candidates, estimate classifier-induced pairwise distinguishability between their propagated receiver-intensity distributions, and select a K-symbol alphabet. The score, derived from held-out pairwise cross-entropy, is an operational measure of empirical separability rather than a source-space metric (Barrett & Myers, 2004; Lopez-Paz & Oquab, 2018). The building blocks are established: a calibrated image classifier (Guo et al., 2017), the equal-prior identity relating binary cross-entropy risk to Jensen–Shannon divergence (Lin, 1991; Goodfellow et al., 2014), farthest-point sampling under chordal projective distance (Conway et al., 2002; Love et al., 2003), and the discrete max–min (p-dispersion) formulation with exact clique solvers (Hochwald & Marzetta, 2000; Erkut, 1990; Martí et al., 2022; Gonzalez, 1985; Tomita & Kameda, 2007). What we develop is the classifier-induced pairwise cross-entropy distinguishability score itself, together with its multi-seed, multi-realization-bank pooling protocol, and the comparison of the two selection objectives it supports.

We instantiate this framework with the $d = 4$ Laguerre–Gaussian basis from the tutorial benchmark of Avramov-Zamurovic et al. (2023), $M = 500$ candidates, and $K = 15$ selected symbols. On propagation realizations disjoint from selection, both selected alphabets outperform the binary reference in all 25 matched evaluation cells; the min-sum alphabet raises mean accuracy from 65.26% to 79.06%. Figure 1 illustrates the pipeline used throughout this work.

We summarize our main contributions as follows.

1. We formulate receiver-distribution alphabet design in random media and make it operational without an analytic channel metric, hand-specified receiver transform, or differentiable channel model, using a classifier-induced pairwise cross-entropy score pooled across encoder seeds and realization banks.
2. We pose two selection objectives on the pooled score and ask which yields the better-decoding alphabet: a max–min rule protecting the least distinguishable selected pair, solved exactly with a global optimality certificate over the candidate pool, and a min-sum rule minimizing the total pairwise loss, solved by multistart greedy search.
3. On propagation realizations disjoint from selection, the min-sum selected alphabet improves mean accuracy by 13.80 percentage points over the binary reference and wins in all

25 matched evaluation cells of the studied propagation regime; it also beats max–min by 4.39 points, winning 22 of the 25 cells. Within this regime and under this score, min-sum is the better of the two, which is a measurement rather than a general claim.

The rest of the paper is organized as follows. Section 2 specifies the physical setting including the incident-mode coefficient space and random-medium propagation model that produce receiver-intensity observations, then formalizes the receiver-distribution alphabet-design problem and constructs the finite candidate pool over which selection is performed. Section 3 builds the classifier-induced pairwise cross-entropy distinguishability score from receiver-intensity images, then poses the two selection objectives defined on that score. Section 4 reports the independent classification performance of the selected alphabets and objective-level stability of the selection criterion. Section 5 concludes and discusses limitations and future directions.

2 Physical Setting and Problem Formulation

A beam is fixed here by a handful of coefficients, but what the receiver records is not: every pass through a random medium distorts the same incident field differently, so what a designer must reason about is the spread of receiver intensities a beam produces rather than any one image. Section 2.1 fixes the incident field and its coefficient space, and Section 2.2 sends it through the random medium to define the receiver intensity. Section 2.3 then attaches a class-conditional observation distribution to each beam class and states alphabet design as keeping those distributions apart. Since that distribution has no closed form in the coefficients, Section 2.4 replaces the continuous coefficient space with a binary reference alphabet and a finite candidate pool built around it.

2.1 Input field and coefficient space

In structured-light communication, each transmitted symbol is a distinct transverse field pattern that the receiver must identify from the intensity image recorded after propagation through the random medium (Avramov-Zamurovic et al., 2023). We build these patterns as superpositions of d fixed basis modes $\psi_1, \dots, \psi_d$. A vector $\alpha = (\alpha_1, \dots, \alpha_d)$ on the unit sphere $S^{d-1} := \{\alpha \in \mathbb{R}^d : \|\alpha\|_2 = 1\}$ defines the incident field

$$u_\alpha(0, x) = c_\alpha \sum_{r=1}^{d} \alpha_r \psi_r(x). \tag{1}$$

Here $x \in \mathbb{R}^2$ is the transverse coordinate, and $c_\alpha > 0$ normalizes every field to the same discrete power $P_{\text{src}} = 10^5$; the explicit normalization is stated in Appendix A.1. Fixing the discrete power with normalization equalizes the transmit energy of every candidate, so that a class cannot be made easier to detect by being brighter and distinguishability reflects spatial structure alone. Real coefficients are a modeling choice rather than a consequence of intensity detection: they fix intermodal relative phases to $\{0, \pi\}$ and keep the space comparable with the binary reference alphabet. Given real coefficients, a global sign is unobservable in intensity, so α and $-\alpha$ represent the same beam class and $\mathcal{X}_d := S^{d-1}/\{\alpha \sim -\alpha\} \cong \mathbb{RP}^{d-1}$, with $[\alpha]$ the class of α. As our benchmark we use the $d = 4$ Laguerre–Gaussian basis of Avramov-Zamurovic et al. (2023), whose modes ψ_r have radial and azimuthal indices $(p_r, \ell_r) \in \{(0, 1), (1, 4), (0, -6), (1, 8)\}$ and whose standard formula is given in Appendix A.1; its coefficient space is $\mathcal{X}_4 \cong \mathbb{RP}^3$.

2.2 Random-medium propagation and receiver measurement

Once the incident field leaves the transmitter, it has to cross the random medium before anything is measured. Refractive-index fluctuations along the path accumulate phase distortion that differs from one realization to the next. Let $u(z, x) \in \mathbb{C}$ be the wave field at propagation distance z. The paraxial model (Andrews et al., 2001; Schmidt, 2010; Fannjiang & Sølna, 2004) is

$$\left(2ik_0\partial_z + \Delta_x + k_0^2\nu(z, x)\right) u(z, x) = 0, \qquad u(0, x) = u_\alpha(0, x), \tag{2}$$

where $k_0 > 0$ is the wavenumber, $\Delta_x := \partial_{x_1}^2 + \partial_{x_2}^2$ is the transverse Laplacian, and ν is the random refractive-index fluctuation.

We represent ν by independent von Kármán phase screens in a split-step Fourier simulation (Andrews et al., 2001; Schmidt, 2010; Bal & Nair, 2025b). The transverse correlation of each screen follows the von Kármán power spectral density:

$$\Phi_{\mathrm{vK}}(\kappa) = \frac{\sigma^2}{16}\,\frac{\exp(-|\kappa|^2/\kappa_\ell^2)}{(|\kappa|^2+\kappa_0^2)^{11/6}}, \qquad \kappa_0 = \frac{2\pi l_0}{L_0}. \tag{3}$$

Here $\kappa \in \mathbb{R}^2$ is transverse spatial frequency, σ controls amplitude, $\kappa_0 = 2\pi l_0/L_0$ is the low-frequency cutoff, l_0 is a transverse reference length, L_0 is the outer-scale length, and κ_ℓ is the high-frequency cutoff. At each propagation step j, a real phase screen χ_ω^j is synthesized from Φ_{vK} by Fourier filtering of complex Gaussian noise, $\chi_\omega^j = \mathrm{Re}\{\mathcal{F}^{-1}[(\Delta z\,\Phi_{\mathrm{vK}})^{1/2}\xi_\omega^j]\}$, where ξ_ω^j is unit complex Gaussian noise on the frequency grid, Δz is the propagation step, and $\mathcal{F}^{-1}$ is the inverse Fourier transform.

At receiver distance $z_\star$, the measured intensity image is $Y_{\alpha,\omega} := |u_\omega(z_\star,\cdot)|^2$. Throughout, we work in a single fixed propagation regime, $\sigma = 3$ and $z_\star = 5$ in the dimensionless units of Appendix A.3. Appendix A.4 gives the reason: its measured scintillation index places the regime beyond the weak-fluctuation band, while boundary leakage stays below 1.5%. Since both the paraxial equation (2) and the discrete split-step operator are linear in the initial field, $u_{-\alpha}(z,\cdot) = -u_\alpha(z,\cdot)$. Consequently, $Y_{-\alpha,\omega} = |-u_\alpha(z_\star,\cdot)|^2 = |u_\alpha(z_\star,\cdot)|^2 = Y_{\alpha,\omega}$. Thus, the receiver measurement depends only on the projective class $[\alpha] \in \mathcal{X}_d$. Appendix A.2 writes out the phase-screen sampling and the discrete split-step operator, and Appendix A.3 lists the physical and numerical parameters.

Although the coefficient space $\mathcal{X}_d$ is finite-dimensional, the receiver observation law induced by each $[\alpha]$ lies in the much larger, generally infinite-dimensional space of probability measures. For each fixed random-medium realization, propagation is linear in the incident field, but intensity detection makes the map $\alpha \mapsto Y_{\alpha,\omega} = |u_\omega|^2$ nonlinear. After marginalizing over the random medium, the resulting distribution has no available closed form in terms of α (Bal & Nair, 2025a). Consequently, source-space geometry alone does not yield a tractable distinguishability objective or its optimizer directly, motivating the problem formulation and the solution development in the receiver-distribution space.

2.3 Receiver-distribution alphabet-design problem

We now formulate the design problem for the coefficient space $\mathcal{X}_d$, selected alphabet size K, and candidate-pool size M. For each beam class $[\alpha] \in \mathcal{X}_d$, let $P_\alpha := \mathrm{Law}(Y_{\alpha,\omega})$ denote the probability distribution of the receiver intensity over random-medium realizations ω. The problem is to select an alphabet $\mathcal{A} \subset \mathcal{X}_d$ with $|\mathcal{A}| = K$ whose observation distributions can be reliably distinguished, in the sense that a classifier reading the receiver intensity rarely confuses them; Section 3 makes this precise.

2.4 Discretization of the candidate pool

As an intuitive choice, we retain the binary source alphabet containing $2^d - 1$ nonzero binary classes as the reference:

$$\mathcal{A}_{\mathrm{bin}} = \left\{ \left[\frac{b}{\|b\|_2} \right] : b \in \{0,1\}^d \setminus \{\mathbf{0}\} \right\}. \tag{4}$$

The underlying problem selects K beam classes from the continuous space $\mathcal{X}_d$. To obtain a tractable discrete approximation, we restrict the feasible classes to a fixed pool $\mathcal{C} = \{[\alpha^{(0)}], \ldots, [\alpha^{(M-1)}]\} \subset \mathcal{X}_d$ with $M \geq 2^d - 1$. The pool contains $\mathcal{A}_{\mathrm{bin}}$ and adds $M - (2^d - 1)$ classes by farthest-point sampling. For projective classes, we use the chordal distance on real projective space, equivalently on the Grassmannian of lines $G(d,1)$ (Conway et al., 2002; Love et al., 2003):

$$d_{\mathrm{proj}}([\alpha],[\beta]) = \sqrt{1 - |\alpha^{\mathsf{T}}\beta|^2}. \tag{5}$$

Explicitly, we first draw N_{res} vectors $g_\ell \sim \mathcal{N}(0, I_d)$, normalize them as $q_\ell = g_\ell/\|g_\ell\|_2$, and form the reservoir $\mathcal{U} = \{[q_\ell] : \ell = 1, \ldots, N_{\mathrm{res}}\}$. Then we start from $\mathcal{C}_0 = \mathcal{A}_{\mathrm{bin}}$. Step t selects

$$[q_t^\star] \in \operatorname*{arg\,max}_{[q]\in\mathcal{U}\setminus\mathcal{C}_t} \min_{[\alpha]\in\mathcal{C}_t} d_{\mathrm{proj}}([q],[\alpha]), \qquad \mathcal{C}_{t+1} = \mathcal{C}_t \cup \{[q_t^\star]\}. \tag{6}$$

Repeating this update for $M-(2^d-1)$ steps spreads the candidates over $\mathcal{X}_d$ while ensuring that all 2^d-1 binary reference classes remain in the pool. Here d_{proj} covers $\mathcal{X}_d$ rather than predicting which classes the receiver can tell apart: it leaves no large region unsampled, while every separability claim is made by the score of Section 3.

The optimization stages select K classes using the distributions $P_{\alpha^{(i)}}$. Hence, every optimality statement is conditional on $\mathcal{C}$, not on the continuous space $\mathcal{X}_d$. Appendix B.1 gives the random seed and index convention. For the experiments reported in this paper, we instantiate this notation with $d=4$ Laguerre–Gaussian basis modes, $K=15$, $M=500$, and $N_{\mathrm{res}}=100{,}000$; hence the binary reference alphabet has $2^4-1=15$ classes and farthest-point sampling adds 485 candidates.

3 Selection on the Pooled Distinguishability Matrix

Section 2 leaves a finite candidate pool with no number attached to any pair in it. This section supplies one and then selects on it. We train a classifier on receiver intensities from the pool, measure how distinguishable a pair is by how often the classifier confuses it, and select the K classes of the alphabet from the resulting scores. Section 3.1 trains the encoder and Section 3.2 turns its confusions into the pooled distinguishability matrix. Two selection objectives are then natural on that matrix, and this section poses both. Section 3.3 solves the max–min problem on that matrix exactly, and Section 3.4 poses the min-sum problem and solves it by multistart greedy search, and Section 4 measures which of them yields the better-decoding alphabet.

3.1 Encoder and training procedure

For candidate $i \in \{0,\dots,M-1\}$, let $Y^A_{i,m}$ denote the receiver intensity generated with random-medium realization m from the training set A. We train a one-channel ResNet-18 classifier without pretrained weights on standardized receiver images $\widetilde{Y}$; Appendix B.2 gives the normalization convention and architecture details. For an input image Y, the learned feature map f_ϕ produces logits and class probabilities

$$s(Y) = W f_\phi(\widetilde{Y}) + b \in \mathbb{R}^M, \qquad \pi_k(Y) = \frac{\exp(s_k(Y))}{\sum_{c=0}^{M-1}\exp(s_c(Y))}, \tag{7}$$

where s_k and π_k are the logit and predicted probability for candidate k with respect to the encoder. Writing $\theta=(\phi,W,b)$, the training objective is the multiclass cross-entropy below:

$$\widehat{L}(\theta) = -\frac{1}{Mn}\sum_{i=0}^{M-1}\sum_{m=1}^{n}\ln \pi_i(Y^A_{i,m}), \tag{8}$$

where m indexes the n independent realizations of candidate i, with $n=n_A=80$ in our experiments. Before pairwise cross-entropy estimation, the frozen encoder logits are temperature-calibrated on a held-out set T with one scalar per encoder seed, which places all seeds on a common logit scale before the losses of Section 3.2 are pooled; see details in Appendix B.3.

3.2 Pairwise cross-entropy distinguishability

We propose to measure the separability of each candidate pair by the empirical binary cross-entropy of the calibrated pairwise posterior induced by the trained encoder. This gives a direct operational score: pairs that the encoder separates with lower held-out cross-entropy are treated as more distinguishable under random-medium propagation.

Let $s_k(Y)$ be logit k from the encoder and let $\widehat{\beta}$ be the inverse temperature fitted on set T. For distinct candidate classes i and j, the calibrated pairwise posterior probability is

$$\widehat{\eta}_{ij}(Y) = \frac{\exp(\widehat{\beta}s_i(Y))}{\exp(\widehat{\beta}s_i(Y))+\exp(\widehat{\beta}s_j(Y))} = \frac{1}{1+\exp\big(-\widehat{\beta}(s_i(Y)-s_j(Y))\big)}. \tag{9}$$

Given an independent realization bank C containing n_C realizations per candidate ($n_C=20$ in our experiments), with each candidate receiving equal prior weight $1/2$, the pairwise cross-entropy loss

between classes i and j is

$$L_{ij} = -\frac{1}{2n_C} \sum_{m=1}^{n_C} \left[\ln \widehat{\eta}_{ij}(Y_{i,m}^C) + \ln\bigl(1 - \widehat{\eta}_{ij}(Y_{j,m}^C)\bigr) \right] \qquad (i \neq j), \tag{10}$$

with diagonal $L_{ii} = 0$. Note that $1 - \widehat{\eta}_{ij}(Y) = \widehat{\eta}_{ji}(Y)$.

To mitigate variations in encoder initialization and random-medium sampling, we average L_{ij} across independent encoder fits and held-out realization banks to obtain the pooled loss matrix $\overline{L}$ (detailed in Appendix B.4). The pooled distinguishability matrix $\overline{D}$ is then defined by

$$\overline{D}_{ij} = \exp(-\overline{L}_{ij}) \qquad (i \neq j), \qquad \overline{D}_{ii} = 0. \tag{11}$$

Because smaller cross-entropy loss indicates stronger pairwise separability, smaller $\overline{L}_{ij}$ corresponds to larger distinguishability $\overline{D}_{ij} \in (0, 1]$.

What we would like to score for a pair of candidates is how far apart their class-conditional observation distributions lie. Let $\mathcal{Y}$ be the receiver-intensity image space; for distinct candidates i and j, write $P_i = P_{\alpha(i)}$ and $P_j = P_{\alpha(j)}$ for their distributions on $\mathcal{Y}$ and $P_{ij}^{\mathrm{mix}} = (P_i + P_j)/2$ for their equal mixture. Their separation is measured by the Jensen–Shannon (JS) divergence (Lin, 1991), built from the Kullback–Leibler (KL) divergence $\mathrm{KL}(P\|Q) = \int_{\mathcal{Y}} \ln(\mathrm{d}P/\mathrm{d}Q)\,\mathrm{d}P$ of measures $P \ll Q$ (Kullback & Leibler, 1951):

$$\mathrm{JS}(P_i, P_j) = \frac{1}{2}\,\mathrm{KL}(P_i\|P_{ij}^{\mathrm{mix}}) + \frac{1}{2}\,\mathrm{KL}(P_j\|P_{ij}^{\mathrm{mix}}). \tag{12}$$

A classical identity connects this divergence to classification. At equal priors, the smallest binary cross-entropy risk on the pair equals $\ln 2 - \mathrm{JS}(P_i, P_j)$. Lower cross-entropy therefore means larger JS divergence.

We cannot evaluate this divergence directly: the densities come from the random-medium simulation and have no closed form. We therefore replace both terms of the identity, the calibrated posterior $\widehat{\eta}_{ij}$ of Equation (9) for the optimal classifier and the realization-bank average of Equation (10) for the population risk, and use $\overline{D}$ as a monotone operational distinguishability score induced by calibrated encoders. Appendix B.4 gives the identity and the lower bound on $\mathrm{JS}(P_i, P_j)$ that any classifier yields through it.

3.3 Exact max–min optimization

Let the candidate index set be $\{0, \ldots, M-1\}$. For any subset $S \subset \{0, \ldots, M-1\}$ with $|S| = K$, define its minimum pairwise distinguishability under $\overline{D}$ by

$$d_{\min}(\overline{D}, S) = \min_{\substack{i,j \in S \\ i<j}} \overline{D}_{ij}. \tag{13}$$

An exact max–min alphabet denoted as S_{mm} maximizes this minimum pairwise distinguishability over all candidate subsets of cardinality K:

$$S_{\mathrm{mm}} \in \operatorname*{arg\,max}_{\substack{S \subset \{0,\ldots,M-1\} \\ |S|=K}} d_{\min}(\overline{D}, S), \qquad \tau^\star = d_{\min}(\overline{D}, S_{\mathrm{mm}}) = \max_{\substack{S \subset \{0,\ldots,M-1\} \\ |S|=K}} d_{\min}(\overline{D}, S). \tag{14}$$

The max–min criterion explicitly protects the least distinguishable selected pair by maximizing the worst-case pairwise separation across all candidate subsets. Formulation (14) is an instance of the discrete p-dispersion problem (Erkut, 1990; Martí et al., 2022). Direct enumeration is out of reach: our $M = 500$, $K = 15$ setting has $\binom{500}{15} \approx 1.89 \times 10^{28}$ candidate subsets. The associated fixed-threshold feasibility problem is NP-complete in general (Karp, 1972). Appendix C.1 gives the threshold formulation, exact solver, and certificate. Conditional on the finite candidate pool and empirical pooled matrix $\overline{D}$, the resulting alphabet S_{mm} is globally optimal for Equation (14).

3.4 The min-sum objective

For any alphabet $S \subset \{0, \ldots, M-1\}$ with $|S| = K$, define the total pairwise cross-entropy loss by

$$L_{\mathrm{sum}}(\overline{L}, S) = \sum_{\substack{i,j \in S \\ i<j}} \overline{L}_{ij}. \tag{15}$$

The max–min objective of Equation (14) optimizes the bottleneck $\tau^\star$ alone and is blind to the remaining $\binom{K}{2} - 1$ pairwise losses. The complementary choice is to optimize their total: it weighs every selected pair, and protects no single one of them. We therefore pose the min-sum selection problem

$$S_{\mathrm{sum}} \in \operatorname*{arg\,min}_{\substack{S \subset \{0,\dots,M-1\} \\ |S|=K}} L_{\mathrm{sum}}(\overline{L}, S). \tag{16}$$

Which of them yields the better-decoding alphabet does not follow from the definitions, so we treat it as an empirical question and answer it in Section 4.1. Problem (16) is a minimum-weight cardinality-K clique problem on the complete graph with edge weights $\overline{L}_{ij}$, which is NP-hard (Karp, 1972). It admits an exact mixed-integer linear programming (MILP) formulation (Nemhauser & Wolsey, 1988), but a global solve is costly, particularly when selection must be repeated across stability checks; see details in Appendix C.3.

We solve Equation (16) by multistart greedy search. One greedy run evaluates every exchange that replaces a selected candidate by an unselected one and accepts the largest strictly positive decrease in L_{sum}, terminating when no exchange improves it by more than $\varepsilon = 10^{-12}$; for N_{ex} accepted exchanges the $N_{\mathrm{ex}}+1$ neighborhood scans cost $\mathcal{O}((N_{\mathrm{ex}}+1)K^2(M-K))$ once the pairwise matrices are formed. We run 256 seeded starts and keep the best local optimum found, which costs $177.87\,\mathrm{s}$ in total. This returns no optimality certificate, and we do not claim one. What we report instead is saturation: across 256 greedy runs and 206 distinct local optima, the best objective is attained early, is never improved afterwards, and is reached from five different starts at the identical alphabet; see details in Appendix C.2. An exact solver does not beat it either: started from that alphabet, Gurobi (Gurobi Optimization, LLC, 2026) neither improves on it within a full hour nor proves it optimal; see details in Appendix C.3. Saturation is evidence that further restarts would not help, not a proof that no better alphabet exists; the alphabet S_{sum} it returns is optimal for Equation (16) only as far as that evidence reaches.

4 Experiments and Results

Everything Section 3 optimizes is conditional on the pooled matrices $\overline{D}$ and $\overline{L}$, which particular encoders read off particular realization banks; a certificate on those matrices says nothing by itself about how the selected beams behave in fresh random media. The experiments therefore test the selection off those matrices. Four alphabets are compared: the two selected ones, S_{mm} and S_{sum}, one for each of the two objectives of Section 3; the binary reference $\mathcal{A}_{\mathrm{bin}}$ of Equation (4); and a geometry-only control S_{geo}, which applies our own exact max–min criterion to the projective distance of Equation (5) between source coefficient vectors, and so asks whether the receiver distributions were needed. Each alphabet is decoded by classifiers retrained from scratch on propagation realizations disjoint from every realization used to fit the encoders or estimate the pairwise scores, in 25 matched evaluation cells (five partition seeds $\times$ five classifier-training seeds); within a cell every alphabet shares the partition, the training seed, and the random media. Section 4.1 reports the decoding results, and Section 4.2 withholds one encoder seed or one realization bank at a time and reruns the whole selection, asking whether the choice was an accident of how the scores were estimated. Appendix E gives the control's construction, optimality certificate, and decoding protocol.

4.1 Decoding selected and unadapted alphabets on fresh media

Table 1 reports the outcome. Both selected alphabets beat the binary reference in every one of the 25 cells, by 9.40 ± 4.03 percentage points (pp) for exact max–min and 13.80 ± 4.62 pp for min-sum. The geometry-only alphabet does not merely fail to match them: it falls 17.95 ± 3.55 pp below exact max–min and 22.34 ± 3.65 pp below min-sum, losing all 25 cells against each, and it is 8.54 ± 4.39 pp below the unadapted binary reference. Also, four alphabets selected from the same candidate pool on source-intensity dissimilarity instead each lose all 25 cells to the binary reference; we defer the details to Appendix F. The two objectives are also separated from each other, and min-sum wins: the paired per-cell difference $S_{\mathrm{sum}} - S_{\mathrm{mm}}$ is 4.39 ± 3.44 pp, winning 22 of the 25 cells (paired $t = 6.38$, $p < 10^{-5}$).

Worst-class recall tells the same story more sharply. Within a cell, the worst-class recall is the smallest per-class recall over the 15 symbols, each decoded from 30 held-out realizations. No class

of either selected alphabet falls below 30% recall in any cell, whereas the geometry-only control and the binary reference both contain a class reduced to a single correct realization in thirty. The weakest class follows the same ordering as accuracy, and not the bottleneck: min-sum holds the lower $d_{\min}(\overline{D}, S)$ of the two selected alphabets yet has the higher worst-class recall. Appendix E.4 compares both with the receiver scores of all four alphabets.

Alphabet	Accuracy (%)	Mean Worst Recall (%)	Min Worst Recall (%)	Wins vs. Binary
Geometry-only control (S_{geo})	56.72 ± 2.68	20.27	3.33	$0/25$
Binary reference	65.26 ± 3.92	26.80	3.33	—
Exact max–min (S_{mm})	74.67 ± 2.90	42.00	30.00	$25/25$
Min-sum (S_{sum})	79.06 ± 4.28	51.60	40.00	$25/25$

Table 1: Independent 15-class classification performance over 25 matched evaluation cells (5 partition seeds $\times 5$ classifier seeds, one classifier fit per alphabet per cell). Accuracy is reported as empirical mean $\pm$ sample standard deviation over cells; recall columns summarize the worst class within each cell. The $\pm$ is the pooled spread across cells, which mixes partition and training-seed variation; all comparisons in the text are within-cell paired differences. Appendix E details the geometry-only control.

Figure 2 shows the selected min-sum alphabet S_{sum} at both ends of the channel. The petal counts and clean rings of panel (a) dissolve into speckle, leaving only coarse envelope structure. What one realization cannot show is the spread of receiver intensities each class produces over the medium, which is what $\overline{D}$ scores.

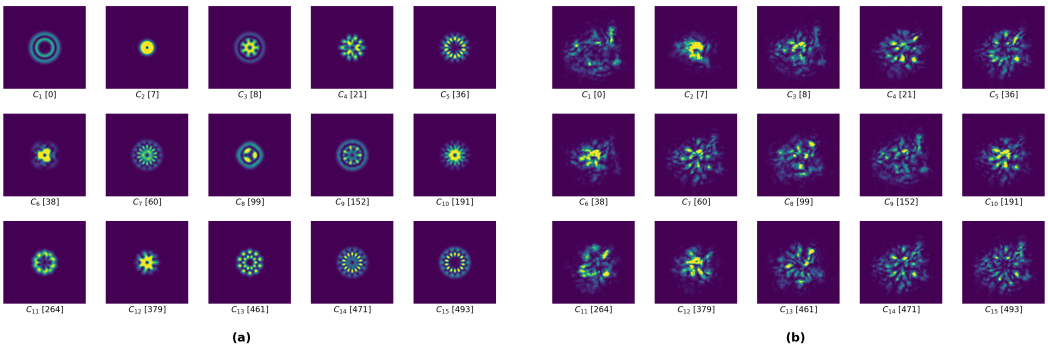

Figure 2: Source and receiver intensities for the fifteen classes of the min-sum alphabet S_{sum} selected in Section 3.4. Panel (a) shows the source intensities $|u_\alpha(0,\cdot)|^2$ from Equation (1), and panel (b) shows one receiver intensity $Y_{\alpha,\omega}$ per class, all generated under the same random-medium realization ω. Labels give the class index and candidate index.

Figure 3 compares the class-level prediction structure of the binary reference and min-sum alphabets over the same 25 matched evaluation cells. The min-sum matrix exhibits stronger diagonal concentration and fewer large off-diagonal confusions, consistent with its higher mean test accuracy in Table 1.

4.2 Objective-level stability of pairwise cross-entropy selection

The pooled matrices rest on 25 pairwise cross-entropy matrices, one per combination of five encoder seeds and five held-out realization banks. We omit one factor at a time, pool the remaining 20, resolve the min-sum problem by multistart greedy search, and score the resulting alphabet on the full pooled $\overline{D}$ and $\overline{L}$.

Table 2 shows that every omission keeps at least 95.82% of the $d_{\min}$ of the full-data min-sum alphabet and increases L_{sum} by at most 8.61%; encoder-seed omissions move both more than realization-bank omissions do. Some candidate identities change under omissions, but the optimized objective values remain stable. Encoder seeds also agree on which pairs are hard: the Spearman rank correlation between the five encoder-averaged matrices has mean 0.915 and minimum 0.897. Appendix D.2 gives the per-omission tables, the spread of individual entries, and a decoding check of the least favorable omission.

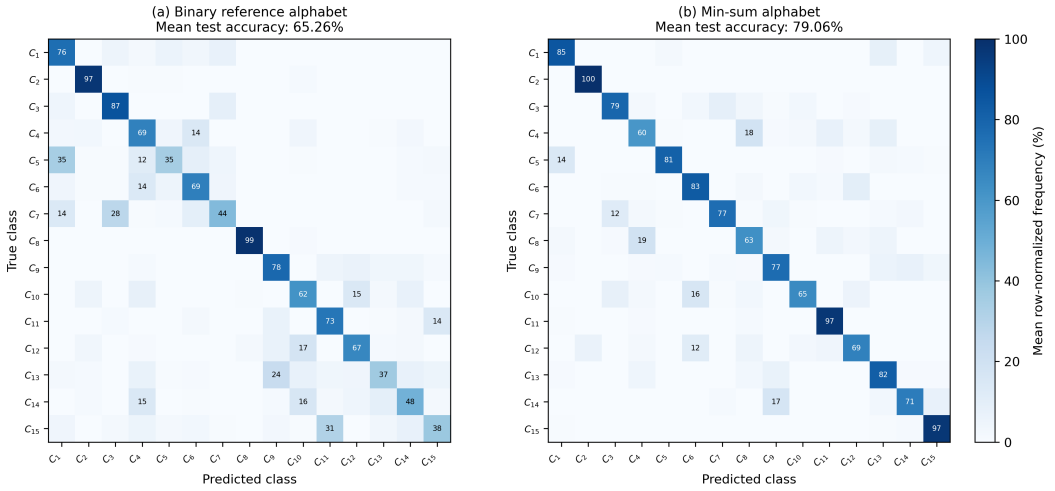

Figure 3: Averaged row-normalized independent-test confusion matrices for (a) the binary reference alphabet and (b) the min-sum alphabet S_{sum}. Each panel averages 25 confusion matrices after row normalization (5 partition seeds $\times$ 5 classifier-training seeds). Rows denote true classes and columns denote predicted classes. Every diagonal value and every off-diagonal value of at least 10% is annotated. Class labels C_1 through C_{15} are local to each alphabet and must not be compared cell by cell across panels.

Omitted Factor	Mean $d_{\min}$ Ret. (%)	Min. $d_{\min}$ Ret. (%)	Mean ΔL_{sum} (%)	Max. ΔL_{sum} Incr. (%)
Encoder seed	98.51	95.82	3.78	8.61
Realization bank	99.86	98.05	2.06	6.94

Table 2: Selected-alphabet objective-level stability under omission of individual encoder seeds and held-out realization banks. Each row summarizes five leave-one-factor-out reselections. The reported $d_{\min}$ retention and relative change in summed pairwise cross-entropy (ΔL_{sum}) are evaluated on the full pooled matrices $\overline{D}$ and $\overline{L}$, relative to the full-data min-sum alphabet.

5 Conclusion

We studied receiver-distribution alphabet design for structured light, where distinguishability is determined by random receiver-intensity images rather than by source-space geometry alone. By replacing the source metric with a classifier-induced pairwise distinguishability score, we obtain alphabets that consistently improve over the unadapted binary reference in independent matched evaluations. Of the two objectives posed on that score, min-sum decodes better in the propagation regime studied. The main limitation is that both the search domain and the distinguishability scores are empirical approximations: discretizing $\mathbb{RP}^3$ to $M = 500$ classes restricts the exact max–min certificate to $\mathcal{C}$ under the pooled matrix $\overline{D}$, rather than the continuous design space or population receiver-image distributions. A second limitation is that the min-sum alphabet carries no optimality certificate, only the saturation evidence of Section 3.4. Future work should therefore develop adaptive or continuous optimization over $\mathbb{RP}^3$ (and $\mathbb{CP}^3$ for complex coefficients), incorporate lower-confidence-bound constraints to account for pairwise-estimation uncertainty, and investigate scalable solvers beyond multistart greedy search: the time-limited MILP of Appendix C.3 neither improves on the min-sum alphabet from multistart greedy search nor closes its bound, so whether a better alphabet exists in the pool remains open.

Reproducibility Statement

The candidate-pool construction, data partitions, random seeds, model training protocol, dissimilarity estimator, selection algorithms, and evaluation statistics are specified in the main text and appendices. An anonymous code archive accompanies this submission as supplementary material. It contains the experiment scripts that produced the reported numbers, the fixed candidate pool and its manifest, the pooled pairwise cross-entropy and distinguishability matrices, the selected alpha-

bets, the per-run predictions and confusion counts, and the recorded configurations, random seeds, and content hashes. The selection results, and the table entries computed from the pooled matrices or from the recorded classification runs, can be re-derived from the archive without re-running propagation; the entries of Tables 4, 5 and 6 are included as computed. The archive's README maps each table and figure to the script and files that produce it. The per-realization receiver-intensity arrays and the trained encoder checkpoints are omitted for size; they are regenerated from the seeds and configurations recorded in Appendices B.1 and B.3.

AI USE STATEMENT

Generative AI tools were used to assist with manuscript organization, LaTeX migration, and language editing. They were not used as evidence for the scientific claims or as a substitute for the reported numerical experiments. The authors reviewed the AI-assisted material and take responsibility for the final text, claims, and artifacts.

REFERENCES

Larry C. Andrews, Ronald L. Phillips, and Cynthia Y. Hopen. *Laser Beam Scintillation with Applications*, volume 99. SPIE Press, 2001.

Svetlana Avramov-Zamurovic, Joel M. Esposito, and Charles Nelson. Classifying beams carrying orbital angular momentum with machine learning: tutorial. *J. Opt. Soc. Am. A*, 40(1):64–77, Jan 2023. doi: 10.1364/JOSAA.474611. URL https://opg.optica.org/josaa/abstract.cfm?URI=josaa-40-1-64.

Guillaume Bal and Anjali Nair. Long distance propagation of wave beams in paraxial regime, 2025a. URL https://arxiv.org/abs/2409.09514.

Guillaume Bal and Anjali Nair. Splitting algorithms for paraxial and Itô–Schrödinger models of wave propagation in random media, 2025b. URL https://arxiv.org/abs/2503.00633.

Harrison H. Barrett and Kyle J. Myers. *Foundations of Image Science*. Wiley-Interscience, New York, 2004. ISBN 978-0-471-15300-9.

Jaime E. Cisternas and Jaime A. Anguita. Selection of OAM signal constellations for atmospheric channels using optimal transport theory. *Opt. Express*, 32(24):43445–43458, Nov 2024. doi: 10.1364/OE.537609. URL https://opg.optica.org/oe/abstract.cfm?URI=oe-32-24-43445.

J. H. Conway, R. H. Hardin, and N. J. A. Sloane. Packing lines, planes, etc.: Packings in Grassmannian space, 2002. URL https://arxiv.org/abs/math/0208004.

Timothy Doster and Abbie T. Watnik. Machine learning approach to OAM beam demultiplexing via convolutional neural networks. *Appl. Opt.*, 56(12):3386–3396, Apr 2017. doi: 10.1364/AO.56.003386. URL https://opg.optica.org/ao/abstract.cfm?URI=ao-56-12-3386.

Erhan Erkut. The discrete p-dispersion problem. *European Journal of Operational Research*, 46(1): 48–60, 1990. ISSN 0377-2217. doi: 10.1016/0377-2217(90)90297-O. URL https://www.sciencedirect.com/science/article/pii/037722179090297O.

Albert C. Fannjiang and Knut Sølna. Scaling limits for beam wave propagation in atmospheric turbulence. *Stochastics and Dynamics*, 4(01):135–151, 2004. doi: 10.1142/S0219493704000973.

Teofilo F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306, 1985. ISSN 0304-3975. doi: 10.1016/0304-3975(85)90224-5. URL https://www.sciencedirect.com/science/article/pii/0304397585902245.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014. URL https://arxiv.org/abs/1406.2661.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural networks, 2017. URL https://arxiv.org/abs/1706.04599.

Gurobi Optimization, LLC. Gurobi Optimizer Reference Manual, 2026. URL https://www.gurobi.com.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. URL https://arxiv.org/abs/1512.03385.

B.M. Hochwald and T.L. Marzetta. Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading. *IEEE Transactions on Information Theory*, 46(2):543–564, 2000. doi: 10.1109/18.825818.

William A. Jarrett, Svetlana Avramov-Zamurovic, Joel M. Esposito, K. Peter Judd, and Charles Nelson. Neural network classification of beams carrying orbital angular momentum after propagating through controlled experimentally generated optical turbulence. *J. Opt. Soc. Am. A*, 41 (6):B1–B13, Jun 2024. doi: 10.1364/JOSAA.515096. URL https://opg.optica.org/josaa/abstract.cfm?URI=josaa-41-6-B1.

T. Kailath. The divergence and Bhattacharyya distance measures in signal selection. *IEEE Transactions on Communication Technology*, 15(1):52–60, 1967. doi: 10.1109/TCOM.1967.1089532.

Richard M. Karp. *Reducibility among Combinatorial Problems*, pp. 85–103. Springer US, Boston, MA, 1972. ISBN 978-1-4684-2001-2. doi: 10.1007/978-1-4684-2001-2_9. URL https://doi.org/10.1007/978-1-4684-2001-2_9.

Mario Krenn, Robert Fickler, Matthias Fink, Johannes Handsteiner, Mehul Malik, Thomas Scheidl, Rupert Ursin, and Anton Zeilinger. Communication with spatially modulated light through turbulent air across Vienna. *New Journal of Physics*, 16(11):113028, November 2014. ISSN 1367-2630. doi: 10.1088/1367-2630/16/11/113028. URL http://dx.doi.org/10.1088/1367-2630/16/11/113028.

S. Kullback and R. A. Leibler. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951. doi: 10.1214/aoms/1177729694. URL https://doi.org/10.1214/aoms/1177729694.

Hsuan-Tien Lin, Chih-Jen Lin, and Ruby C. Weng. A note on Platt's probabilistic outputs for support vector machines. *Machine Learning*, 68(3):267–276, 2007. doi: 10.1007/s10994-007-5018-6.

J. Lin. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151, 1991. doi: 10.1109/18.61115.

Sanjaya Lohani and Ryan T. Glasser. Turbulence correction with artificial neural networks. *Opt. Lett.*, 43(11):2611–2614, Jun 2018. doi: 10.1364/OL.43.002611. URL https://opg.optica.org/ol/abstract.cfm?URI=ol-43-11-2611.

David Lopez-Paz and Maxime Oquab. Revisiting classifier two-sample tests, 2018. URL https://arxiv.org/abs/1610.06545.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL https://api.semanticscholar.org/CorpusID:53592270.

D.J. Love, R.W. Heath, and T. Strohmer. Grassmannian beamforming for multiple-input multiple-output wireless systems. *IEEE Transactions on Information Theory*, 49(10):2735–2747, 2003. doi: 10.1109/TIT.2003.817466.

Rafael Martí, Anna Martínez-Gavara, Sergio Pérez-Peló, and Jesús Sánchez-Oro. A review on discrete diversity and dispersion maximization from an OR perspective. *European Journal of Operational Research*, 299(3):795–813, 2022. ISSN 0377-2217. doi: 10.1016/j.ejor.2021.07.044. URL https://www.sciencedirect.com/science/article/pii/S0377221721006548.

George L. Nemhauser and Laurence A. Wolsey. *Integer and Combinatorial Optimization*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, New York, 1988. URL https://api.semanticscholar.org/CorpusID:42076848.

Timothy O'Shea and Jakob Hoydis. An introduction to deep learning for the physical layer. *IEEE Transactions on Cognitive Communications and Networking*, 3(4):563–575, 2017. doi: 10.1109/TCCN.2017.2758370.

John C. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In Alexander J. Smola, Peter Bartlett, Bernhard Schölkopf, and Dale Schuurmans (eds.), *Advances in Large Margin Classifiers*, pp. 61–74. MIT Press, 1999.

Jason D. Schmidt. *Numerical Simulation of Optical Wave Propagation with Examples in MATLAB*. SPIE, 2010. doi: 10.1117/3.866274.

Maximilian Stark, Faycal Ait Aoudia, and Jakob Hoydis. Joint learning of geometric and probabilistic constellation shaping. In *2019 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, 2019. doi: 10.1109/GCWkshps45667.2019.9024567.

Etsuji Tomita and Toshikatsu Kameda. An efficient branch-and-bound algorithm for finding a maximum clique with computational experiments. *Journal of Global Optimization*, 37(1):95–111, 2007. doi: 10.1007/s10898-006-9039-7. URL https://api.semanticscholar.org/CorpusID:27578449.

Aokun Wang, Anjali Nair, Zhongjian Wang, and Guillaume Bal. ML-based approach to the classification and generation of structured light propagation in turbulent media. *J. Opt. Soc. Am. A*, 43 (7):C61–C70, Jul 2026. doi: 10.1364/JOSAA.600455. URL https://opg.optica.org/josaa/abstract.cfm?URI=josaa-43-7-C61.

A Physical and Numerical Model

This appendix records the incident-mode formula, power normalization, random phase-screen synthesis, split-step numerical propagation and parameter values omitted from Section 2. All alphabet comparisons use this same propagation condition.

A.1 Incident modes and power normalization

Write the transverse coordinate as $x = (\varrho\cos\varphi, \varrho\sin\varphi)$. The Laguerre–Gaussian source mode with radial index $p \in \mathbb{N}_0$, azimuthal index $\ell \in \mathbb{Z}$, and beam waist $w_0 > 0$ is

$$\psi_{p,\ell}(\varrho,\varphi) = \sqrt{\frac{2p!}{\pi(p+|\ell|)!}}\,\frac{1}{w_0}\left(\frac{\sqrt{2}\varrho}{w_0}\right)^{|\ell|} L_p^{(|\ell|)}\left(\frac{2\varrho^2}{w_0^2}\right)\exp\left(-\frac{\varrho^2}{w_0^2}+i\ell\varphi\right), \tag{17}$$

where $L_p^{(|\ell|)}$ is the generalized Laguerre polynomial. On the transverse grid $\mathcal{G}$ with spacing Δx, the normalization in Equation (1) is

$$c_\alpha = \left[\frac{P_{\mathrm{src}}}{(\Delta x)^2\sum_{x\in\mathcal{G}}\left|\sum_{r=1}^{4}\alpha_r\psi_r(x)\right|^2}\right]^{1/2}, \qquad P_{\mathrm{src}} = 10^5. \tag{18}$$

This discrete normalization fixes the transmitted power on the same numerical grid used for propagation. It prevents the selection procedure from exploiting trivial total-power differences and makes the comparison depend on the resulting receiver-intensity distributions.

A.2 Random phase-screen synthesis and split-step propagation

This subsection records both the random phase-screen sampling and the split-step receiver measurement. The simulation represents ν by independent phase screens along z (Fannjiang & Sølna, 2004;

Bal & Nair, 2025b). The power law and finite low- and high-frequency cutoffs in Equation (3) represent outer- and inner-scale effects for atmospheric turbulence over the resolved spatial frequencies (Andrews et al., 2001; Schmidt, 2010). At step j for random-medium realization ω, for each component $a \in \{1, 2\}$ and discrete frequency κ, we draw

$$\xi^j_{a,\omega}(\kappa) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \qquad a \in \{1,2\}. \tag{19}$$

Here $\mathcal{N}(0,1)$ is the standard normal distribution, and the samples are independent over a, ω, j, and κ. Define $\xi^j_\omega(\kappa) := \xi^j_{1,\omega}(\kappa) + i\xi^j_{2,\omega}(\kappa)$. The Fourier-domain screen is

$$\widehat{\chi}^j_\omega(\kappa) = \sqrt{\Delta z\, \Phi_{\mathrm{vK}}(\kappa)}\, \xi^j_\omega(\kappa), \tag{20}$$

where Δz is the propagation step. The real screen is

$$\chi^j_\omega = \mathrm{Re}\{\mathcal{F}^{-1}\left[\widehat{\chi}^j_\omega\right]\}. \tag{21}$$

Here $\mathcal{F}^{-1}$ is the inverse Fourier transform. Thus, screens are independent across j, while Φ_{vK} determines their transverse correlation.

Split-step propagation and receiver measurement. Writing u^j_ω for the discrete transverse field after j propagation steps of size Δz, the split-step Fourier update is

$$u^{j+1}_\omega = \mathcal{F}^{-1}\Big[e^{-i\Delta z|\kappa|^2/(2k_0)}\mathcal{F}\Big\{e^{i\chi^j_\omega}u^j_\omega\Big\}\Big], \qquad u^0_\omega = u_\alpha(0,\cdot), \tag{22}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the two-dimensional discrete Fourier transform and its inverse.

A.3 Physical and numerical parameters

Table 3 gives the setting used for every reported alphabet comparison. The split-step Fourier computation uses periodic boundary conditions on the transverse domain. This boundary convention is the numerical closure associated with the Fourier propagation step; it is held fixed across all candidate classes and selected alphabets. The von Kármán parameters in the table define the single turbulence spectrum used throughout the experiments, including the low- and high-frequency cutoffs. These

Table 3: Physical and numerical parameters for random-medium propagation.

Symbol	Meaning	Value
L_x	transverse domain side length	64
N_x	grid points per transverse dimension	256
Δx	transverse grid spacing	$1/4$
$z_\star$	receiver distance	5
Δz	propagation-step size	$1/32$
k_0	dimensionless wavenumber	$1/2$
w_0	incident-beam waist	4
σ	spectrum-amplitude parameter	3
l_0	transverse reference length	1.5
L_0	outer-scale spectrum parameter	256
κ_ℓ	inner-scale high-frequency cutoff	5.92

values give $\Delta\kappa = \pi/32$, $\kappa_0 = 3\pi/256$, and $J = z_\star/\Delta z = 160$.

Two conventions in this construction should be stated, because they determine what the parameters of Table 3 mean numerically. First, the discrete spectrum is mapped to the screen with the frequency-grid differential applied explicitly: the sampled amplitude carries a factor $2\pi/\Delta\kappa$ and the inverse transform carries $\Delta\kappa^2/(4\pi^2)$, so the $1/N_x^2$ normalization of the discrete inverse transform cancels and the variance of each screen is independent of N_x at fixed L_x. The complex Gaussian draw has unit variance in each of its real and imaginary parts, and taking the real part retains exactly half of it; no further factor is absorbed into σ. Under these conventions the per-step phase variance, written $\mathrm{Var}\,\chi^j_\omega$, is the variance of one screen at one point with its spatially constant $\kappa = 0$ term excluded, because that term cancels in $|u|^2$. It equals $\Delta z\,\Delta\kappa^2(4\pi^2)^{-1}\sum_{\kappa\neq 0}\Phi_{\mathrm{vK}}(\kappa)$, which at the parameters of Table 3 is 0.118 rad^2. Here and below, RMS (root-mean-square) denotes the square

root of a mean of squared values. The screen phase has zero mean, so its RMS is the square root of the per-step phase variance: the per-step RMS phase is 0.344 rad.

Second, no low-frequency subharmonic sampling is used. The grid resolves no transverse frequency below $\Delta\kappa = \pi/32$, whereas the spectrum places its low-frequency rollover at $\kappa_0 = 3\pi/256$, a factor $8/3$ lower. The interval $(0, \Delta\kappa)$ containing the rollover is therefore not sampled, and the effective outer scale of the realized medium is bounded by the transverse domain $L_x = 64$ rather than by L_0. For this reason L_0 is listed above as a spectrum parameter: it fixes κ_0 and hence the shape of the sampled spectrum, but the screens do not realize an outer scale of 256. The practical consequence is that the screens act as a high-pass spatial filter, and that macroscopic wavefront tilt is under-represented relative to a fully sampled von Kármán medium. All candidate classes, selected alphabets and baselines are propagated through this identical numerical channel, so the comparisons reported in this paper are unaffected; what the convention limits is the correspondence between L_0 and any physical outer scale.

A.4 Measured diagnostics of the propagation regime

Table 4 reports diagnostics measured from recorded receiver intensities over a sweep of propagation regimes. At each regime the binary reference alphabet is propagated for 500 random-medium realizations per class. Medium seeds are matched across σ: the amplitude parameter enters the spectrum only as a multiplier, so a shared seed gives the same underlying Gaussian field rescaled, and the σ rows are therefore a paired comparison rather than independent draws. The accumulated RMS phase is $\sqrt{J \operatorname{Var} \chi_\omega^j}$, the RMS of the phase that all J screens add along the path; unlike the other columns, it is computed from the spectrum rather than measured. The scintillation index is the variance of receiver intensity divided by its squared mean, estimated per pixel over realizations and averaged over a per-class circular aperture whose radius is fixed once at $\sigma = 3$, $z_\star = 5$ to enclose 90% of the mean power and then held identical at every regime; an aperture recomputed per regime would move with the quantity being swept. The beam width is the power-weighted RMS radius of one receiver-intensity image about its centroid, averaged over realizations. The relative wander is the RMS distance of that centroid from its mean position over realizations, divided by the beam width. The speckle correlation length is the first radius, in pixels, at which the radially averaged periodic autocorrelation of the mean-subtracted receiver-intensity image falls below e^{-1}, averaged over realizations. Boundary leakage is the fraction of receiver power in a strip of width $L_x/10$ along each edge of the transverse domain.

Table 4: Diagnostics measured from recorded receiver intensities across propagation regimes, for the binary reference alphabet at 500 realizations per class. The accumulated RMS phase is computed from the spectrum, not measured. Values are means over the 15 classes. One quantity is swept per block, with the others held at the values shown. The step size is $\Delta z = 1/32$ throughout except in the last block, which repeats the strongest swept regime, $\sigma = 4.5$ at $z_\star = 5$, with $\Delta z = 1/128$ to control for discretization. The row in bold is the regime used for every alphabet comparison in this paper.

Swept setting	Value	accumulated RMS phase	scintillation index	relative wander	speckle corr. length (px)	boundary leakage
Turbulence strength σ held $z_\star = 5$	0.50	0.73	0.250	0.034	36.1	0.03%
	0.75	1.09	0.405	0.051	35.0	0.07%
	1.00	1.45	0.558	0.067	33.3	0.12%
	1.50	2.18	0.849	0.098	29.0	0.29%
	2.00	2.90	1.118	0.125	24.4	0.55%
	3.00	**4.35**	**1.589**	**0.171**	**18.0**	**1.47%**
	4.50	6.53	2.073	0.214	13.6	4.22%
Receiver distance $z_\star$ held $\sigma = 3$	2.5	3.08	1.420	0.098	23.9	0.04%
	10.0	6.15	1.985	0.162	10.5	25.25%
Step size Δz	$1/128$	6.53	2.055	0.212	13.6	4.11%

Two measurements bracket the setting used throughout. From below, the scintillation index exceeds one from $\sigma \approx 1.78$ and is 1.59 at the reported regime, with its worst class at 2.41. In the weak-fluctuation band it stays below one (Andrews et al., 2001), so the reported regime lies beyond that

band. From above, the periodic boundary condition of the split-step operator binds: at $z_\star = 10$ a quarter of the mean receiver power, and 35% for the worst class, has reached that edge strip, where the Fourier propagation step wraps it back into the field. At $z_\star = 5$ that fraction is 1.47%. The domain cannot simply be enlarged to relieve this, because L_x also sets the effective outer scale of the sampled spectrum: without subharmonic sampling, doubling L_x to 128 more than doubles the per-step phase variance, changing the medium rather than only the numerics.

These two bounds delimit a window rather than a point: any σ between roughly 1.78 and 4.5 satisfies both, and $\sigma = 3$ is interior to it. Within that window the reported regime is the more demanding choice at each coordinate. Of the two admissible distances, $z_\star = 5$ carries the higher scintillation index and the finer speckle, 1.59 and 18.0 pixels against 1.42 and 23.9 at $z_\star = 2.5$, while keeping leakage below 1.5%. In the other direction $\sigma = 4.5$, the only stronger amplitude swept, nearly triples the leakage to 4.22% and raises the per-step RMS phase past 0.5 rad, so it sits at the edge of the window rather than inside it.

The per-step phase variance also fixes how much of the incident field survives in the mean. The screens are independent Gaussian fields, and each split-step update preserves power. Removing each screen's spatially constant phase leaves every receiver intensity unchanged. After that removal, the mean receiver field, the average of the complex field at $z_\star$ over random-medium realizations, is the freely propagated incident field attenuated by $\exp(-J \operatorname{Var} \chi^j_\omega/2)$. The fraction of receiver power it carries, the mean-field fraction, is $\exp(-J \operatorname{Var} \chi^j_\omega)$: the exponential of minus the squared accumulated RMS phase. The kinetic limit of Bal & Nair (2025a) is the scaling in which this exponent is of order one and the mean receiver field survives. The mean-field fraction is 0.59 at $\sigma = 0.5$ and 0.12 at $\sigma = 1$, but 6×10^{-9} at the reported regime, where the accumulated RMS phase is 4.35 rad. The reported regime therefore lies past the kinetic limit.

The last row controls the step size. Refinement in Δz is the check Bal & Nair (2025b) apply to their split-step schemes, which they show can converge even when Δz is not the smallest scale of the problem. At $\sigma = 4.5$, the largest swept amplitude and the one with the largest per-step RMS phase, refining Δz from $1/32$ to $1/128$ changes the scintillation index by 0.87%, the beam width by 0.11%, and the speckle correlation length by 0.29%. The step size of Table 3 is therefore converged across the sweep, and a fortiori at $\sigma = 3$, where the per-step RMS phase is smaller.

B Candidate Pool, Data Partitions, and Encoder Training Protocol

This appendix records the fixed candidate-pool construction, data partitions, encoder architecture, training protocol, calibration procedure, and pooled pairwise cross-entropy estimator used to construct $\overline{L}$ and $\overline{D}$.

B.1 Candidate-pool construction

The reservoir in Section 2.4 is generated with seed 224,879. If several reservoir classes attain the same value in Equation (6), the class with the smallest reservoir index is selected. The resulting pool uses indices $0, \dots, 14$ for the binary reference classes and indices $15, \dots, 499$ for the added classes. The candidate-pool size $M = 500$ and the reservoir size $N_{\mathrm{res}} = 100{,}000$ of Section 2.4 are compute-limited design choices rather than tuned values. M fixes the encoder's output dimension and the number of candidate pairs $\binom{M}{2} = 124{,}750$.

B.2 Input normalization and network architecture

Input standardization and split isolation. Each receiver observation is a single-channel scalar intensity image on the transverse grid $\Omega = \{1, \dots, N_x\}^2$, with $N_x = 256$. We standardize every image using the empirical mean μ_A and population-standard-deviation convention σ_A computed over all pixels, candidates, and realizations in the encoder-training set A:

$$\mu_A = \frac{1}{M n_A |\Omega|} \sum_{i,m,u} Y^A_{i,m}(u), \qquad \sigma_A^2 = \frac{1}{M n_A |\Omega|} \sum_{i,m,u} \big(Y^A_{i,m}(u) - \mu_A\big)^2, \tag{23}$$

where the sums range over $\{i = 0, \ldots, M-1\}$, $\{m = 1, \ldots, n_A\}$, and $\{u \in \Omega\}$. The normalized input is $\widetilde{Y} = (Y - \mu_A)/\sigma_A \in \mathbb{R}^{1\times 256\times 256}$. The statistics μ_A and σ_A are frozen after the encoder-training split is prepared and are reused without recomputation for validation, temperature calibration, pairwise cross-entropy estimation, and independent evaluation. Thus, no validation or test image contributes to the normalization constants.

The encoder uses the standard ResNet-18 basic-block configuration (He et al., 2015) with stage depths $(2, 2, 2, 2)$, but it is trained from scratch rather than initialized from ImageNet weights. This matches the scalar receiver-intensity input and avoids importing a three-channel natural-image stem. The first layer is a one-channel 7×7 convolution with stride 2 and padding 3, followed by batch normalization, ReLU, and 3×3 max pooling with stride 2 and padding 1. The four residual stages use channel widths $(64, 128, 256, 512)$, with spatial downsampling at the usual stage transitions. For a 256×256 input, the final convolutional feature map has 512 channels on an 8×8 grid, and global average pooling produces $h(\widetilde{Y}) \in \mathbb{R}^{512}$.

A learned affine projection maps this pooled feature to the 128-dimensional embedding $f_\phi(\widetilde{Y})$, followed by the jointly trained 500-class affine head in Equation (7). No additional activation, normalization layer, or dropout is inserted between $h(\widetilde{Y})$, $f_\phi(\widetilde{Y})$, and the logits $s(Y)$. In the notation of Section 3.1, ϕ includes the convolutional stem, residual stages, global pooling, and the 512-to-128 affine projection; W and b are the parameters of the final candidate-class head.

B.3 Training, checkpoint selection, and calibration

For each encoder seed, set A contains $Mn_A = 500 \cdot 80 = 40{,}000$ training images. Training runs for 200 epochs, with 250 balanced mini-batches per epoch. Each mini-batch samples 16 distinct candidate classes and 4 realizations from each sampled class, giving batch size 64; classes and realizations may reappear in later mini-batches.

We optimize Equation (8) with AdamW (Loshchilov & Hutter, 2019), fixed learning rate 10^{-4}, and weight decay 10^{-3}. The AdamW moment parameters are $(\beta_1, \beta_2) = (0.9, 0.999)$ and $\epsilon = 10^{-8}$, with no learning-rate schedule. For epoch e, the balanced sampler is initialized deterministically from e, the encoder-training seed r, n_A, and the fixed random-medium generation seed 7,739,241. If $\theta_{r,t}$ denotes the raw encoder parameters after update t for encoder seed r, the exponential moving average (EMA) parameters are updated with decay 0.999:

$$\overline{\theta}_{r,0} = \theta_{r,0}, \qquad \overline{\theta}_{r,t} = 0.999\,\overline{\theta}_{r,t-1} + 0.001\,\theta_{r,t}. \tag{24}$$

After every epoch, the EMA checkpoint is evaluated on all $Mn_B = 500 \cdot 20 = 10{,}000$ set-B images. We retain the EMA checkpoint with the lowest set-B cross-entropy and freeze that checkpoint before temperature calibration and pairwise cross-entropy estimation. Because pairwise cross-entropy uses log probabilities, logit calibration prevents scale discrepancies across encoder seeds. For each encoder seed r with logits $s^{(r)}(Y)$, we fit one scalar inverse temperature β following Guo et al. (2017) on an independent held-out set T, which contains $n_T = 30$ realizations per candidate ($Mn_T = 500 \cdot 30 = 15{,}000$ images in total). For $\beta \geq 0$, the temperature-scaled predictive distribution for candidate k is

$$\pi^{(r)}_{\beta,k}(Y) = \frac{\exp(\beta s^{(r)}_k(Y))}{\sum_{c=0}^{M-1} \exp(\beta s^{(r)}_c(Y))}. \tag{25}$$

We fit the inverse temperature $\widehat{\beta}_r$ by minimizing multiclass cross-entropy on set T:

$$\widehat{\beta}_r \in \operatorname*{arg\,min}_{0\leq\beta\leq 100} -\frac{1}{Mn_T} \sum_{i=0}^{M-1} \sum_{m=1}^{n_T} \ln \pi^{(r)}_{\beta,i}(Y^T_{i,m}). \tag{26}$$

The minimization uses bounded scalar search with absolute tolerance 10^{-10}. Only this scalar is fitted on set T; the encoder and classification head remain frozen.

Because the classification head is linear, rescaling (W, b) reproduces Equation (25) exactly, so $\widehat{\beta}_r$ adds no capacity the head does not already have. What it adds is an out-of-sample fit: the head's scale minimizes Equation (8) on set A, and the weight decay of 10^{-3} penalizes that scale, whereas $\widehat{\beta}_r$ minimizes Equation (26) on the disjoint set T with no penalty. The fitted values in Table 5 show the two optima do not coincide.

Table 5: Fitted inverse temperatures and their effect on held-out negative log-likelihood (NLL) on set T, $Mn_T = 15{,}000$ images per encoder seed. Every $\widehat{\beta}_r < 1$, so the fit softens a systematically overconfident head; held-out NLL falls by 0.093–0.398 nats. The spread of $\widehat{\beta}_r$ across seeds is a factor of 1.33, and it is the scale discrepancy that Equation (36) would otherwise inherit.

Encoder seed r	$\widehat{\beta}_r$	$1/\widehat{\beta}_r$	NLL at $\beta = 1$	NLL at $\widehat{\beta}_r$
42	0.6247	1.601	4.1422	3.9134
100	0.6243	1.602	4.1056	3.8744
2025	0.7332	1.364	3.8384	3.7457
31415	0.6633	1.508	4.0424	3.8770
271828	0.5528	1.809	4.3250	3.9272

B.4 Pairwise cross-entropy estimation and pooling

This subsection specifies the information-theoretic foundation of the pairwise cross-entropy score, the multi-seed pairwise cross-entropy estimates, and the pooling step that produces the empirical matrices $\overline{L}$ and $\overline{D}$ used for alphabet selection.

Information-theoretic foundation and variational Jensen–Shannon bound. We detail here the connection between binary cross-entropy risk and the symmetric Jensen–Shannon (JS) divergence (Lin, 1991) summarized in Section 3.2. For distinct candidate classes i and j, let P_i and P_j be their receiver-image distributions on $\mathcal{Y}$ and let $P_{ij}^{\mathrm{mix}} = (P_i + P_j)/2$ be their equal mixture. For any measurable binary classifier $\eta : \mathcal{Y} \to [0, 1]$, the equal-prior binary cross-entropy (CE) risk is defined by

$$\mathcal{L}_{ij}(\eta) = -\frac{1}{2}\mathbb{E}_{Y\sim P_i}[\ln \eta(Y)] - \frac{1}{2}\mathbb{E}_{Y\sim P_j}[\ln(1 - \eta(Y))]. \tag{27}$$

The equal-prior binary-classification identity (Lin, 1991; Goodfellow et al., 2014) establishes that

$$\mathrm{JS}(P_i, P_j) = \ln 2 - \inf_{\eta:\mathcal{Y}\to[0,1]} \mathcal{L}_{ij}(\eta), \tag{28}$$

where the infimum is attained, up to P_{ij}^{mix}-null sets, by the Bayes-optimal classifier

$$\eta_{ij}^{\star}(Y) = \frac{\mathrm{d}P_i(Y)}{\mathrm{d}P_i(Y) + \mathrm{d}P_j(Y)}. \tag{29}$$

Consequently, any learned classifier η provides a variational lower bound on the true Jensen–Shannon divergence:

$$\ln 2 - \mathcal{L}_{ij}(\eta) \le \mathrm{JS}(P_i, P_j). \tag{30}$$

In the continuous optical propagation setting, the likelihood ratio $\mathrm{d}P_i/\mathrm{d}P_j$ cannot be computed analytically due to multi-slice random phase fluctuations in the random medium. The calibrated neural posterior $\widehat{\eta}_{ij}$ in Equation (9) serves as a parameterized surrogate for $\eta_{ij}^{\star}$, and evaluating binary cross-entropy over finite realization banks approximates the population risk $\mathcal{L}_{ij}(\widehat{\eta}_{ij})$ via Monte Carlo sampling. While the rigorous lower-bound interpretation (30) holds strictly for the population risk $\mathcal{L}_{ij}$, the empirical distinguishability matrix $\overline{D}_{ij} = \exp(-\overline{L}_{ij})$ acts as an operational separability score that monotonically ranks candidate pairs according to their distinguishability under the learned encoder.

Random seed sets and realization banks. Encoder training uses a set of five independent initialization seeds:

$$\mathcal{S} = \{42, 100, 2025, 31415, 271828\}. \tag{31}$$

For each encoder seed $r \in \mathcal{S}$, the ResNet-18 encoder and classification head are trained on set A ($n_A = 80$ realizations per candidate) and the checkpoint with lowest loss on set B ($n_B = 20$ realizations per candidate) is selected, followed by fitting the inverse temperature $\widehat{\beta}_r$ on set T ($n_T = 30$ realizations per candidate).

Pairwise cross-entropy evaluation is performed over a set of five independent held-out realization banks:

$$\mathcal{Q} = \{101, 211, 307, 401, 503\}. \tag{32}$$

Each realization bank C_q ($q \in \mathcal{Q}$) contains $n_C = 20$ random-medium realizations per candidate class. Within a fixed realization bank C_q, realization $m \in \{1, \ldots, n_C\}$ uses the same random medium across all M candidates. This common random number design evaluates pairwise comparisons under matched propagation conditions and reduces Monte Carlo variation.

Calibrated posterior and pairwise loss matrix. For encoder seed $r \in \mathcal{S}$ with logit function $s^{(r)}(Y)$ and fitted inverse temperature $\widehat{\beta}_r$, the calibrated pairwise posterior probability for classes i and j is

$$\widehat{\eta}_{ij}^{(r)}(Y) = \frac{\exp(\widehat{\beta}_r s_i^{(r)}(Y))}{\exp(\widehat{\beta}_r s_i^{(r)}(Y)) + \exp(\widehat{\beta}_r s_j^{(r)}(Y))} = \frac{1}{1 + \exp\big(-\widehat{\beta}_r (s_i^{(r)}(Y) - s_j^{(r)}(Y))\big)}. \tag{33}$$

The inverse temperature $\widehat{\beta}_r$ is fitted once per encoder seed against the full M-class softmax of Equation (26) and is then applied to isolated pairwise margins, so $\widehat{\eta}_{ij}^{(r)}$ is an amortized operational surrogate rather than a posterior calibrated for the binary problem (i, j). Its role is the one stated in Appendix B.3: to standardize logit scale across encoder seeds before the losses are pooled, so that no seed dominates $\overline{L}$ through an arbitrary scale. Where $\widehat{\eta}_{ij}^{(r)}$ departs from $\eta_{ij}^{\star}$, the variational bound (30) is loosened but not invalidated, since it holds for any measurable classifier; this is why $\overline{D}$ is used throughout as a monotone operational score and not as an estimate of the divergence. How large that departure actually is, and whether it moves the ordering, is measured in Table 6 below. On realization bank C_q, with equal prior weights $1/2$ assigned to each candidate, the symmetric pairwise cross-entropy loss is

$$L_{ij}^{(r,q)} = -\frac{1}{2n_C} \sum_{m=1}^{n_C} \Big[\ln \widehat{\eta}_{ij}^{(r)}(Y_{i,m}^{C_q}) + \ln\Big(1 - \widehat{\eta}_{ij}^{(r)}(Y_{j,m}^{C_q})\Big) \Big] \qquad (i \neq j), \tag{34}$$

with diagonal $L_{ii}^{(r,q)} = 0$ and m denoting the realization index. Note that $1 - \widehat{\eta}_{ij}^{(r)}(Y) = \widehat{\eta}_{ji}^{(r)}(Y)$.

Pooling into the selection matrices. For each encoder seed $r \in \mathcal{S}$, the loss matrix averaged over all realization banks $q \in \mathcal{Q}$ is

$$\overline{L}_{ij}^{(r)} = \frac{1}{|\mathcal{Q}|} \sum_{q \in \mathcal{Q}} L_{ij}^{(r,q)}. \tag{35}$$

The fully pooled pairwise cross-entropy matrix $\overline{L}$ averages across all $|\mathcal{S}||\mathcal{Q}| = 25$ seed combinations:

$$\overline{L}_{ij} = \frac{1}{|\mathcal{S}||\mathcal{Q}|} \sum_{r \in \mathcal{S}} \sum_{q \in \mathcal{Q}} L_{ij}^{(r,q)} = \frac{1}{|\mathcal{S}|} \sum_{r \in \mathcal{S}} \overline{L}_{ij}^{(r)}. \tag{36}$$

The pooled distinguishability matrix $\overline{D}$ is obtained by exponentiating the pooled cross-entropy:

$$\overline{D}_{ij} = \exp(-\overline{L}_{ij}) \qquad (i \neq j), \qquad \overline{D}_{ii} = 0. \tag{37}$$

Crucially, exponentiation occurs strictly after pooling the cross-entropy losses over all seeds, $\overline{D}_{ij} = \exp(-\overline{L}_{ij})$, rather than averaging already exponentiated distinguishability matrices.

Measured pairwise calibration and rank stability. The two temperatures are fitted for different purposes, and the diagnostic turns on that difference. $\widehat{\beta}_r$ exists because $\overline{L}$ averages across encoder seeds: pairwise cross-entropy moves with the scale of the logits, the five fitted values in Table 5 span a factor of 1.33, and one scalar per seed keeps that spread out of Equation (36). It is fitted against the full M-class softmax and claims nothing about an individual pair. A per-pair temperature $\beta_{ij}^{(r)}$ is fitted for the opposite thing: freeing $\widehat{\beta}_r$ in Equation (33) and minimizing Equation (34) over $0 \leq \beta \leq 100$ on the $2n_T = 60$ set-T points that carry the pair gives the temperature that pair would have received had it been calibrated for itself. Neither purpose implies the other. Selection, however, consumes neither posterior but the ordering of $\overline{D}$, so the question is whether the two calibrations disagree about that ordering. Table 6 answers it, alongside the binary reliability of $\widehat{\eta}_{ij}^{(r)}$ on C_q. On the worst decile, the tenth of pairs on which bottleneck selection binds, the refitted ordering agrees with the published $\overline{D}$ at $r_s = 0.9620$ for the unsmoothed fit and 0.9684 once its

hard labels are replaced by Platt's smoothed targets (Platt, 1999; Lin et al., 2007). Pool-wide the Platt fit holds at 0.9573, while the unsmoothed fit falls to 0.7066: the 60 points separate 50.2–58.1% of pairs outright, and its minimizer then runs to the bound instead of returning a temperature. That saturation sits on pairs the encoder already tells apart and does not reach the decile selection reads. Rank stability therefore holds under both calibrations, over the part of the ordering that selection uses.

Table 6: Pairwise calibration of $\widehat{\eta}_{ij}^{(r)}$ and rank stability of the pooled ordering. ECE and Brier are pooled over the 1.2475×10^8 binary posteriors on C_q; the third column restricts ECE to the 105 pairs inside the min-sum alphabet S_{sum}. The Spearman correlation r_s is against the published $\overline{D}$, over all $124{,}750$ candidate pairs and over the decile of them with the largest $\overline{L}$, on which selection binds. Every per-pair temperature is fitted on set T and is therefore out of sample on C_q, exactly as $\widehat{\beta}_r$ is.

Inverse temperature	ECE	ECE, S_{sum}	Brier	r_s	r_s, worst decile
None ($\beta = 1$)	0.0105	0.0014	0.0240	—	—
Global $\widehat{\beta}_r$ (published)	0.0029	0.0071	0.0235	1.0000	1.0000
Per-pair β_{ij}, unsmoothed	0.0026	0.0010	0.0229	0.7066	0.9620
Per-pair β_{ij}, Platt targets	0.0391	0.0384	0.0258	0.9573	0.9684

C Optimization Formulations and Certificates

This appendix records the optimization formulations and certificates behind the two selected alphabets. All statements in this appendix are conditional on the fixed $M = 500$ candidate pool and the fixed pooled empirical matrices $\overline{D}$ and $\overline{L}$.

C.1 Exact max–min formulation and optimality proof

For any threshold τ, define the compatibility graph $G_\tau = (V, E_\tau)$ with $V = \{0, \dots, M-1\}$ and

$$E_\tau = \{(i,j) : 0 \le i < j < M,\ \overline{D}_{ij} \ge \tau\}.$$

There exists a size-K subset S with $d_{\min}(\overline{D}, S) \ge \tau$ if and only if G_τ contains a K-clique. Because $d_{\min}(\overline{D}, S)$ is attained by one selected pair, the optimal bottleneck value $\tau^\star$ must be one of the at most $\binom{M}{2}$ off-diagonal entries of $\overline{D}$. The exact max–min problem can therefore be solved by threshold search over the observed off-diagonal values of $\overline{D}$, using an exact clique decision routine at each threshold. The threshold feasibility predicate is monotone: if a threshold is feasible, every lower threshold is feasible; if a threshold is infeasible, every higher threshold is infeasible.

The exact max–min solver used in this paper performs binary threshold search over the sorted unique off-diagonal entries of $\overline{D}$. At each visited threshold τ, feasibility is decided by an exact bitset branch-and-bound clique search with greedy coloring upper bounds (Tomita & Kameda, 2007). The routine either returns a size-K clique witness or certifies that no such clique exists. This dedicated combinatorial solver is used for the exact max–min stage. The global optimality certificate consists of completed feasible or infeasible decisions for the thresholds visited during binary search, together with an infeasibility decision for the threshold immediately above the optimum. For the pooled pairwise-CE distinguishability matrix, all 17 threshold decisions completed. The largest feasible threshold is

$$\tau^\star = 0.9414295324498403,$$

and the next larger observed threshold, 0.9414299915207488, is infeasible for a 15-clique. Hence no 15-candidate subset of the finite pool attains a larger $d_{\min}(\overline{D}, S)$ than $\tau^\star$. The selected exact max–min alphabet is

$$S_{\mathrm{mm}} = \{0, 5, 7, 100, 146, 152, 169, 191, 212, 224, 278, 379, 429, 434, 448\}.$$

It has $d_{\min}(\overline{D}, S_{\mathrm{mm}}) = 0.9414295324498403$ and $L_{\mathrm{sum}}(\overline{L}, S_{\mathrm{mm}}) = 1.7476518732009765$.

C.2 MULTISTART GREEDY SEARCH AND ITS SATURATION

This subsection gives the formal step definitions, tie-breaking rule, and saturation evidence for the multistart greedy search introduced in Section 3.4. One greedy run is defined below. Multistart performs 256 of them from independently drawn starting alphabets and keeps the best local optimum reached. At step $t \geq 0$, given the current alphabet $S^{(t)}$, a candidate exchange replaces one selected candidate class $a \in S^{(t)}$ with one unselected candidate class $b \in \{0, \ldots, M-1\} \setminus S^{(t)}$:

$$S_{\mathrm{can}}(a, b) = \left(S^{(t)} \setminus \{a\}\right) \cup \{b\}. \tag{38}$$

Let F_t denote the set of all such exchange pairs at step t:

$$F_t = S^{(t)} \times \left(\{0, \ldots, M-1\} \setminus S^{(t)}\right). \tag{39}$$

For any exchange pair (a, b), the reduction in summed pairwise cross-entropy loss from $S^{(t)}$ to $S_{\mathrm{can}}(a, b)$ is

$$\Delta_t(a, b) = L_{\mathrm{sum}}(\overline{L}, S^{(t)}) - L_{\mathrm{sum}}(\overline{L}, S_{\mathrm{can}}(a, b)). \tag{40}$$

The best exchange $(a_t^\star, b_t^\star)$ and its corresponding improvement $\Delta_t^\star$ are defined by

$$(a_t^\star, b_t^\star) \in \operatorname*{arg\,max}_{(a,b) \in F_t} \Delta_t(a, b), \qquad \Delta_t^\star = \Delta_t(a_t^\star, b_t^\star) = \max_{(a,b) \in F_t} \Delta_t(a, b), \tag{41}$$

with ties broken deterministically by selecting the smallest removed candidate index a and then the smallest added candidate index b. If $\Delta_t^\star > \varepsilon$, with $\varepsilon = 10^{-12}$, the exchange is accepted by setting $S^{(t+1)} = S_{\mathrm{can}}(a_t^\star, b_t^\star)$, and the search advances to step $t+1$. Otherwise, the search terminates and outputs the local optimum $S^{(t)}$. Multistart returns the best of the 256 local optima so obtained, which is S_{sum}.

Each neighborhood scan considers all $K(M-K)$ candidate exchanges. Given an alphabet S, replacing $a \in S$ by $b \notin S$ replaces exactly $K-1$ pair contributions, so the objective delta compares only the removed (a, j) and added (b, j) terms for $j \in S \setminus \{a\}$. Evaluating each candidate exchange thus scales as $\mathcal{O}(K)$, so each full scan costs $\mathcal{O}(K^2(M-K))$. For N_{ex} accepted moves plus the terminal scan, the total post-matrix selection cost is $\mathcal{O}((N_{\mathrm{ex}}+1)K^2(M-K))$. The number of accepted exchanges N_{ex} is data-dependent; because strict descent prevents revisiting any subset, we have the loose finite bound $N_{\mathrm{ex}} \leq \binom{M}{K} - 1$. One greedy run is therefore polynomial per scan and output-sensitive in N_{ex}, without a polynomial worst-case bound on the overall trajectory. Storage is $\mathcal{O}(M^2)$, dominated by the precomputed matrices $\overline{D}$ and $\overline{L}$. For $M = 500$ and $K = 15$ there are 7,275 exchanges per scan.

The shipped alphabet is

$$S_{\mathrm{sum}} = \{0, 7, 8, 21, 36, 38, 60, 99, 152, 191, 264, 379, 461, 471, 493\},$$

with $L_{\mathrm{sum}}(\overline{L}, S_{\mathrm{sum}}) = 1.4879908200991314$ and $d_{\min}(\overline{D}, S_{\mathrm{sum}}) = 0.9087874529825636$.

Multistart carries no optimality certificate. What is reported in its place is saturation. Figure 4 plots the local optimum of each of the 256 greedy runs and their best-so-far curve, which does not improve after the start at which it first attains 1.4879908200991314. The runs reach 206 distinct local optima, the worst at 1.8569, so they do spread widely; five of them nevertheless return the best objective, and all five return the same fifteen candidate indices. This is evidence that further restarts would not improve the result.

The best value is first attained at start 17 of 256, leaving 239 further runs with no improvement, and recurs at starts 106, 142, 155 and 223. The shipped start count of 256 is justified by this record and was not tuned against the results reported in Section 4.

C.3 EXACT MILP FORMULATION AND RUNTIME COMPARISON

This subsection gives the exact mixed-integer linear programming (MILP) formulation corresponding to Equation (16) and compares it with the multistart greedy search used in the main experiments. A MILP is a linear optimization model in which some decision variables are restricted to be

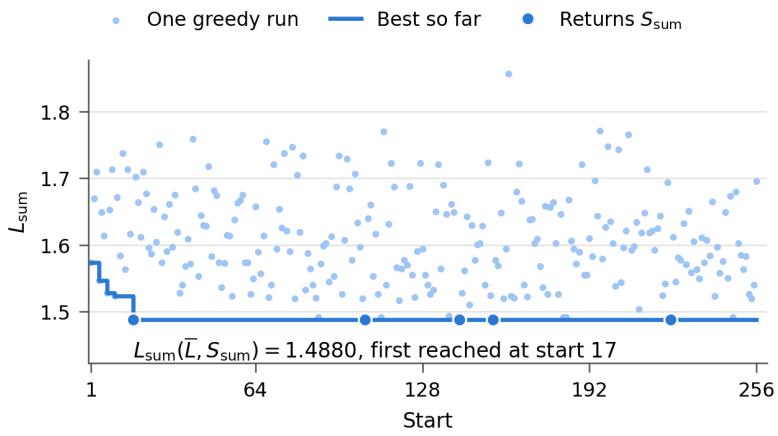

Figure 4: Multistart saturation over the 256 greedy runs of multistart greedy search. Each dot is the local optimum L_{sum} that one greedy run returns; the line is the best objective found after the first n runs; filled markers on the line are the runs that return S_{sum}.

integer-valued; here the integer variables are binary indicators for selected candidates (Nemhauser & Wolsey, 1988). Let $x_i \in \{0,1\}$ indicate whether candidate $i \in \{0,\dots,M-1\}$ is selected.

For each pair $0 \le i < j < M$, introduce continuous variables $y_{ij} \in [0,1]$ that represent the selected-pair product in the objective. Because $\overline{L}_{ij} \ge 0$ and the objective is minimized, the lower linking constraint $y_{ij} \ge x_i + x_j - 1$ is sufficient: selected pairs are charged in the objective, while unselected pairs are driven to zero. The exact MILP is

$$
\begin{aligned}
\min_{x,y} \quad & \sum_{0 \le i < j < M} \overline{L}_{ij} y_{ij} \\
\text{s.t.} \quad & \sum_{i=0}^{M-1} x_i = K, \\
& y_{ij} \ge x_i + x_j - 1, && 0 \le i < j < M, \\
& x_i \in \{0,1\}, && i = 0,\dots,M-1, \\
& 0 \le y_{ij} \le 1, && 0 \le i < j < M.
\end{aligned}
$$

For general M and K, this formulation contains M binary variables, $\binom{M}{2}$ continuous variables, and $1 + \binom{M}{2}$ cardinality and pairwise constraints, apart from variable bounds. Thus, constructing the formulation requires $\Theta(M^2)$ pair data. The combinatorial difficulty is substantially greater: there are $\binom{M}{K}$ candidate assignments satisfying the cardinality constraint, and direct enumeration with objective checks over the $\binom{K}{2}$ selected pairs costs $\Theta\left(\binom{M}{K}\binom{K}{2}\right)$. MILP branch-and-bound can prune many assignments using relaxations and incumbents, but the embedded cardinality-K clique problem is NP-hard, so exact solution retains exponential worst-case complexity (Karp, 1972; Nemhauser & Wolsey, 1988).

The runtime comparison uses the same authoritative pooled matrices $\overline{D}$ and $\overline{L}$, $M = 500$, $K = 15$. Gurobi 13.0.3 (Gurobi Optimization, LLC, 2026) was run with 8 threads, fixed seed 0, and a 3600 s solver time limit. The min-sum MILP was run twice: once cold, with no incumbent supplied, and once warm-started from the min-sum alphabet as a mixed-integer programming start (MIP start). A MIP start is an initial feasible incumbent, i.e., a candidate solution supplied to the solver before optimization begins. The Gurobi route runtime reported in Table 7 includes model construction and optimization, but excludes matrix loading, license authentication, and output writing. The multistart runtime is the total over all 256 greedy runs. For the Gurobi rows, the reported gap is the relative optimality gap $(z_{\text{inc}} - z_{\text{bd}})/|z_{\text{inc}}|$, where z_{inc} is the incumbent objective and z_{bd} is the solver lower bound for this minimization problem.

Table 7: Solver routes on the pooled pairwise-CE min-sum problem, Equation (16). Only the min-sum alphabet is decoded; the time-limited MILP incumbents are not evaluated downstream and no accuracy is claimed for them.

Method	L_{sum}	$d_{\min}(\overline{D})$	Runtime (s)	Gap
Multistart greedy, 256 starts	1.4880	0.9088	177.87	—
Gurobi MILP, cold	1.5931	0.8805	3602.92	0.9917
Gurobi MILP, warm-started	1.4880	0.9088	4041.19	0.9919

Three readings follow. First, the cold MILP is beaten outright by $177.87\,\mathrm{s}$ of multistart greedy search: it terminates at 1.5931 against 1.4880, so its output is a time-limited incumbent and nothing more. Second, warm-started from the multistart incumbent the solver holds that incumbent unchanged for the full hour, explores $643{,}243$ nodes, and returns the identical fifteen candidate indices. The exact route therefore neither improves on multistart nor certifies it. Third, that missing certificate is no evidence against S_{sum}: the gap is set by the bound, and this formulation holds the bound near 0 whatever the incumbent. Table 7 shows this directly, since the cold incumbent is 0.105 worse than S_{sum} yet carries almost the same gap, 0.9917 against 0.9919. The cause is structural rather than a matter of tuning. Setting every x_i to K/M makes the right-hand side $x_i + x_j - 1$ of every lower linking constraint negative, so every y_{ij} can be 0 and the root linear relaxation is identically 0. The valid equality $\sum y_{ij} = \binom{K}{2}$ raises it only to 0.003726, the sum of the $\binom{K}{2} = 105$ smallest pairwise losses in the whole candidate pool, whether or not those pairs lie in one alphabet. This is the cost of a global solve that Section 3.4 anticipates, and it is why S_{sum} rests on the saturation evidence of Appendix C.2 rather than on a solver bound.

D Stability and Supporting Results

This appendix reports the objective-level stability diagnostics that support Section 4.2: matrix stability and leave-one-factor-out reselection.

D.1 Matrix diagnostic and stability metric definitions

Let $E = \{(i,j) : 0 \le i < j < M\}$ be the set of upper-triangular candidate pairs, with $|E| = M(M-1)/2 = 124{,}750$. For matrix stability, we compute one encoder-level dispersion value for each candidate pair: the sample standard deviation of that pair's encoder-averaged cross-entropy across the five encoder-training seeds,

$$s_{\overline{L}}(i,j) = \left[\frac{1}{|\mathcal{S}| - 1} \sum_{r \in \mathcal{S}} \left(\overline{L}_{ij}^{(r)} - \overline{L}_{ij} \right)^2 \right]^{1/2}. \tag{42}$$

Rank consistency is measured by the Spearman correlation between $\mathrm{vec}_E(\overline{L}^{(r)})$ and $\mathrm{vec}_E(\overline{L}^{(r')})$ for each of the ten encoder pairs $r \neq r'$.

For leave-one-factor-out stability, omitting encoder seed r gives

$$\overline{L}_{ij}^{(-r)} = \frac{1}{|\mathcal{S}| - 1} \sum_{r' \in \mathcal{S} \setminus \{r\}} \overline{L}_{ij}^{(r')}, \tag{43}$$

while omitting realization bank q gives

$$\overline{L}_{ij}^{(-q)} = \frac{1}{|\mathcal{S}|(|\mathcal{Q}| - 1)} \sum_{r \in \mathcal{S}} \sum_{q' \in \mathcal{Q} \setminus \{q\}} L_{ij}^{(r,q')}. \tag{44}$$

In both cases, $\overline{D}_{ij}^{(-f)} = \exp(-\overline{L}_{ij}^{(-f)})$ for the omitted factor f. Let S_{-f} denote the alphabet reselected from this reduced 20-matrix pool, and let $S \in \{S_{\mathrm{mm}}, S_{\mathrm{sum}}\}$ be the corresponding full-data alphabet. We report $|S_{-f} \cap S|$, the full-matrix retention ratio $d_{\min}(\overline{D}, S_{-f}) / d_{\min}(\overline{D}, S)$, and the full-matrix relative CE-sum change $(L_{\mathrm{sum}}(\overline{L}, S_{-f}) - L_{\mathrm{sum}}(\overline{L}, S)) / L_{\mathrm{sum}}(\overline{L}, S)$.

D.2 COMPLETE STABILITY RESULTS

Secondary matrix diagnostics. Across all 124,750 upper-triangular candidate pairs of the five encoder-averaged matrices $\overline{L}^{(r)}$ ($r \in \mathcal{S}$), the resulting $s_{\overline{L}}(i,j)$ values have a mean of 0.0216, a median of 0.0093, a 90th percentile (q_{90}) of 0.0556, and a 99th percentile (q_{99}) of 0.1668. Pairwise Spearman rank correlation across all $\binom{5}{2} = 10$ distinct encoder pairs has a mean of 0.9154, a sample standard deviation of 0.0149, a minimum of 0.8966 (pair 42–100), and a maximum of 0.9420 (pair 42–31415), confirming that relative pairwise distinguishability orderings remain consistent across encoder initializations.

Per-omission leave-one-factor-out stability. Tables 8 and 9 report the complete per-omission metrics for both the exact max–min alphabet and the min-sum alphabet across all five encoder omissions and five realization-bank omissions. For exact max–min selection, the relative CE-sum change is a descriptive metric evaluating total cross-entropy on the full matrix $\overline{L}$ under an alphabet chosen purely to maximize bottleneck distinguishability, rather than being an optimized objective.

Table 8: Leave-one-encoder-out selection metrics ($K = 15$, 20-matrix pool per omission) evaluated against the full 25-matrix pooled matrices $\overline{D}$ and $\overline{L}$. Retained candidates are relative to the corresponding full-data alphabet (S_{mm} or S_{sum}). The $d_{\min}$ retention is relative to that of the full-data alphabet (0.9414 for exact max–min, 0.9088 for min-sum).

Seed	Method	Retained	$d_{\min}$ ret. (%)	ΔL_{sum} (%)
42	Exact max–min	14/15	100.00	−1.73
42	Min-sum	13/15	97.27	+0.20
100	Exact max–min	4/15	98.70	+6.67
100	Min-sum	9/15	96.84	+8.61
2025	Exact max–min	2/15	98.94	+8.75
2025	Min-sum	7/15	95.82	+3.75
31415	Exact max–min	4/15	99.56	+5.28
31415	Min-sum	10/15	101.30	+2.55
271828	Exact max–min	12/15	99.40	+2.85
271828	Min-sum	5/15	101.33	+3.81

Table 9: Leave-one-realization-bank-out selection metrics ($K = 15$, 20-matrix pool per omission) evaluated against the full 25-matrix pooled matrices $\overline{D}$ and $\overline{L}$. Retained candidates are relative to the corresponding full-data alphabet (S_{mm} or S_{sum}). The $d_{\min}$ retention is relative to that of the full-data alphabet (0.9414 for exact max–min, 0.9088 for min-sum).

Seed	Method	Retained	$d_{\min}$ ret. (%)	ΔL_{sum} (%)
101	Exact max–min	4/15	99.83	+3.86
101	Min-sum	15/15	100.00	0.00
211	Exact max–min	1/15	99.37	+5.36
211	Min-sum	3/15	102.77	+6.94
307	Exact max–min	14/15	100.00	−1.73
307	Min-sum	11/15	100.33	+0.34
401	Exact max–min	11/15	99.36	+6.10
401	Min-sum	11/15	98.05	+1.08
503	Exact max–min	15/15	100.00	0.00
503	Min-sum	13/15	98.17	+1.94

Decoding the least favorable omission. Objective-level stability is weaker than the claim one would want, so we also decode one omission. Which one was fixed before any omission alphabet existed, by the rule *take the omission retaining the fewest candidates from the full-data alphabet*, ties broken by lower $d_{\min}$ — a rule that selects the *least* favorable arm rather than a convenient one. It is the realization-bank-211 omission, whose alphabet shares just 3 of its 15 symbols with S_{sum} (Table 9). Decoded in the same 25 matched cells of Section 4.1 it reaches $78.73 \pm 3.93\%$ against $79.06 \pm 4.28\%$, a paired difference of -0.33 ± 3.30 pp (winning 11 of 25 cells, $p = 0.62$), and it still beats the binary reference by 13.47 ± 5.09 pp in all 25. Largely different selected identities therefore decode alike. This is one pre-registered omission and not a robustness result: it says the

decoding performance does not hinge on which realization bank entered $\overline{L}$, and says nothing about other propagation regimes.

E GEOMETRY-ONLY ALPHABET CONTROL

This appendix documents the geometry-only control reported in Table 1: how the alphabet S_{geo} was selected, in what sense its selection is optimal, how it was decoded, and why it loses.

E.1 SELECTION CRITERION

The control holds the max–min objective of Equation (14), the candidate pool $\mathcal{C}$, and the alphabet size $K = 15$ fixed, and replaces only the pooled distinguishability matrix $\overline{D}$ by the projective distance d_{proj} of Equation (5) between source coefficient vectors. Selection and tie-breaking see coefficient geometry and candidate ordering only; no receiver intensity and no encoder score enters either. The search uses the same exact clique branch-and-bound solver that produces S_{mm} in Section 3.3, so the comparison isolates the selection criterion rather than the search procedure.

E.2 OPTIMALITY CERTIFICATE

Feasibility is monotone in the threshold, so the optimum is located by binary search over the $124{,}654$ distinct off-diagonal values of the projective distance matrix. Two searches totaling $18{,}000\,\mathrm{s}$ resolved the optimum to a two-sided bracket rather than a single value:

$$\tau^{\star}_{\mathrm{geo}} \in [0.7537082111014741,\ 0.7537124976552428)\,, \tag{45}$$

where the lower endpoint is attained by an exhibited 15-candidate witness and the upper endpoint is proved infeasible by a completed search, leaving exactly one threshold inside the bracket unresolved. The bracket has relative width 5.69×10^{-6}, so $d^{\min}_{\mathrm{proj}}(S_{\mathrm{geo}}) = 0.753708$ — reported as a proved lower bound on $\tau^{\star}_{\mathrm{geo}}$, not as the optimum itself — is within six parts per million of the largest achievable in $\mathcal{C}$. The margin matters only because it forecloses one objection: the control's deficit below the binary reference cannot be attributed to having optimized source-space geometry less carefully than receiver distinguishability.

E.3 DECODING PROTOCOL

S_{geo} was frozen before any decoding, then propagated over the same recorded bank of 100 random-medium realizations, under the same per-partition 50/20/30 train/validation/test assignment, as every other alphabet in Table 1. One decoder was trained from scratch in each of the 25 matched cells, matching the existing architecture, training budget, normalization, model-selection rule, and random-stream conventions, with no alphabet-specific tuning. The other three rows are the previously computed cells, reused after verifying that realizations, partitions, training configuration, candidate mappings, initial-model hashes, and propagation configuration all matched; adding the control required 25 new classifier fits and left the published 75 unchanged.

E.4 SOURCE-SPACE SEPARATION AT THE RECEIVER

Table 10 scores all four alphabets under the pooled receiver matrices $\overline{D}$ and $\overline{L}$, which the geometry-only selection never consulted.

Alphabet	$d_{\min}(\overline{D}, S)$	Mean $\overline{D}_{ij}$	$\sum_{i<j} \overline{L}_{ij}$	Accuracy (%)
Geometry-only (S_{geo})	0.6252	0.9564	5.0979	56.72
Binary reference	0.6544	0.9647	4.0736	65.26
Exact max–min (S_{mm})	0.9414	0.9837	1.7477	74.67
Min-sum (S_{sum})	0.9088	0.9861	1.4880	79.06

Table 10: The four alphabets of Table 1 scored under the pooled receiver matrices, beside their independent decoding accuracy. The mean column averages $\overline{D}_{ij}$ over the $\binom{15}{2}$ selected pairs.

The reason is not that S_{geo} optimized a poor criterion. By Appendix E.1 it maximized the worst pair, exactly as Section 3.3 does, but in source coefficient space. The map from that space to the receiver — random-medium propagation followed by square-law detection — is not an isometry, so candidates that are far apart at the source need not be far apart at the receiver, and the separation S_{geo} secures upstream does not survive the channel.

Two pool-wide statistics show that this is a property of the map rather than of the fifteen candidates S_{geo} happens to contain. Across all $\binom{500}{2} = 124{,}750$ candidate pairs, the Spearman rank correlation between d_{proj} and $\overline{D}$ is 0.24, so source-space separation carries some information about receiver separability but little. Restricting to the top decile of pairs by d_{proj}, whose coefficient vectors are nearly orthogonal, the mean $\overline{D}_{ij}$ is 0.953 and the median 0.988 — yet the minimum is 0.310, essentially the smallest value anywhere in the pool (0.308). Maximal separation in $\mathbb{RP}^3$ does not by itself secure separation at the receiver. Since $\overline{D} = \exp(-\overline{L})$ is strictly decreasing, the corresponding correlation against $\overline{L}$ is the negative of the one quoted.

Across the four alphabets, decoding follows the scores that weigh every selected pair, not the one that weighs only the worst. Both the mean of $\overline{D}_{ij}$ and L_{sum} order the alphabets exactly as their accuracy does, and exactly as their mean worst-class recall in Table 1 does. The bottleneck places both selected alphabets far above the other two but reverses the order between them: S_{sum} has the lower $d_{\min}$, 0.9088 against 0.9414, yet decodes 4.39 pp higher and has the higher mean worst-class recall, 51.60% against 42.00%. Whether or not that difference in $d_{\min}$ exceeds the estimation noise in $\overline{D}$, it cannot account for an accuracy gap that runs against it. Each selected alphabet wins the score it was built to optimize, so what the experiment tests is which score accuracy agrees with, and it agrees with L_{sum}.

E.5 Conclusion

The natural geometric analogue of our own criterion fails: S_{geo} comes within six parts per million of the largest worst-pair projective distance achievable in $\mathcal{C}$, yet decodes 8.54 ± 4.39 pp below the binary reference. Three limits apply. The comparison is conditional on the fixed 500-candidate pool and on the single propagation distance $z_\star = 5$, at the turbulence strength of Appendix A.3. The control tests one source-space rule, max–min projective distance, and a different geometric criterion might order the alphabets differently, so the failure is shown for this rule and not for every source-space criterion. Appendix F tests a second rule of a different kind, on source intensity images rather than coefficient vectors, and its alphabets also decode below the binary reference. Finally, the certificate of Equation (45) is conditional on the recorded projective distance matrix and the fixed pool, not on the continuous coefficient space.

F Source-Intensity-Diversity Alphabet Control

The geometry-only control of Appendix E separates coefficient vectors. Work that designs the alphabet itself has instead emphasized diversity among source intensity profiles (Jarrett et al., 2024). This appendix reports a control of that second kind: alphabets selected for dissimilar clean source intensity images from the candidate pool $\mathcal{C}$ that we construct in Section 2.4, the same $M = 500$ candidates from which every other alphabet in this paper is selected.

F.1 Selection criteria

For candidate i, let $I_i := |u_{\alpha^{(i)}}(0, \cdot)|^2$ be its source intensity: the incident field of Equation (1), power-normalized by Equation (18) on the propagation grid of Table 3, with no random medium drawn and nothing propagated. Two source-intensity dissimilarities are formed over the 256×256 grid pixels:

$$\delta^{\mathrm{L2}}_{ij} = \|I_i - I_j\|_2, \qquad \delta^{\mathrm{corr}}_{ij} = 1 - \mathrm{corr}(I_i, I_j), \tag{46}$$

where corr is the Pearson correlation. With the source power fixed, δ^{L2} counts a concentrated beam as far from a spread one before any spatial comparison. δ^{corr} is invariant to an affine rescaling of either image and registers only whether bright regions coincide. Over the $124{,}750$ candidate pairs the two rank alike, with Spearman correlation 0.904, but their sets of the 1% most dissimilar pairs have a Jaccard index of only 0.455. A max–min objective reads only that tail, so both are used.

Each dissimilarity is paired with two objectives, which gives four alphabets. The source-intensity max–min alphabet solves Equation (14) with $\overline{D}$ replaced by δ, using the threshold search and exact clique solver that produce S_{geo} in Appendix E.1. Both of its certificates are exact: the optimum is attained by an exhibited witness and the next larger distinct value of δ is proved infeasible, so no threshold is left unresolved as in Equation (45). The source-intensity max-sum alphabet instead maximizes $\sum_{i,j\in S,\, i<j} \delta_{ij}$, the source-side counterpart of Equation (16). It is found by the greedy run of Appendix C.2 applied to $-\delta$ from 256 seeded uniform random starts, keeping the best of the 7 (δ^{L2}) and 8 (δ^{corr}) distinct local optima reached, and it carries no certificate. That all four alphabets would be decoded was fixed before any of them was selected, so none was chosen for report after its accuracy was known.

F.2 Decoding

The four alphabets were decoded under the protocol of Appendix E.3, in the same 25 matched cells, with the published cells reused unchanged; the control added 100 classifier fits. Table 11 reports them beside the geometry-only control and the binary reference, together with the pooled receiver scores of Table 10, which none of the four selections consulted.

Alphabet	$d_{\min}(\overline{D}, S)$	Mean $\overline{D}_{ij}$	$\sum_{i<j} \overline{L}_{ij}$	Accuracy (%)	Worst recall (%)	vs. binary (pp)
δ^{L2}, max–min	0.5408	0.9505	5.8265	58.15 ± 2.42	22.13	-7.11 ± 3.82
δ^{corr}, max–min	0.4481	0.9242	9.2413	50.17 ± 2.16	13.87	-15.09 ± 3.38
δ^{L2}, max-sum	0.4249	0.8742	16.8013	39.00 ± 2.35	4.93	-26.26 ± 3.25
δ^{corr}, max-sum	0.3899	0.8490	20.4627	36.54 ± 2.16	4.80	-28.72 ± 3.81
Geometry-only (S_{geo})	0.6252	0.9564	5.0979	56.72 ± 2.68	20.27	-8.54 ± 4.39
Binary reference	0.6544	0.9647	4.0736	65.26 ± 3.92	26.80	—

Table 11: The four source-intensity-diversity alphabets, named by dissimilarity and objective, decoded in the 25 matched cells of Table 1. Receiver scores are under the pooled matrices, as in Table 10. Worst recall is the mean over cells of the worst-class recall. The last column is the paired per-cell difference from the binary reference, mean $\pm$ sample standard deviation over cells; each of the four loses all 25 cells.

Every source-intensity-diversity alphabet loses all 25 cells to the binary reference. The two max–min alphabets fall on either side of the geometry-only control: δ^{L2} decodes 1.43 ± 3.31 pp above it (16 wins, 1 tie, 8 losses; paired $t = 2.16$, $p = 0.041$), and δ^{corr} decodes 6.55 ± 2.43 pp below it, in all 25 cells. Neither shares any of its 15 candidates with S_{geo}, the binary reference, S_{mm} or S_{sum}. The image-space and coefficient-space rules therefore select disjoint alphabets, and both land below the binary reference.

The two max-sum alphabets fail for a reason visible before propagation. A sum can trade a few near-duplicate pairs for many far-apart ones, and here it does. Their median within-alphabet δ^{corr} is 0.66 and 0.77, against 0.42 over the pool, yet they contain 13 and 20 pairs whose source intensities correlate above 0.9; the max–min alphabets contain none. Both hold candidates 141 and 451, whose correlation of 0.992 makes them the 17th most similar of the 124,750 pairs. These two rows therefore measure what a summed source dissimilarity selects more than they measure the channel, and the max–min rows are the comparison with S_{geo}.

F.3 Source intensity and receiver separability

Orthogonal incident fields need not have dissimilar intensities. The Laguerre–Gaussian basis is orthonormal on the propagation grid, so $|\alpha^{\mathsf{T}}\beta| = \sqrt{1 - d_{\mathrm{proj}}([\alpha],[\beta])^2}$ is the overlap of two incident fields, and S_{geo} makes the fields as close to orthogonal as the pool allows. Intensity is quadratic in the field, so two orthogonal fields can still light the same pixels: among the 15,776 pool pairs whose fields overlap by less than 0.1, the source intensities have median correlation 0.47, and up to 0.985.

Source-intensity separation predicts receiver separability somewhat better than coefficient separation does, but still weakly. Over all pool pairs, the Spearman correlation with $\overline{D}$ is 0.38 for δ^{L2} and 0.33 for δ^{corr}, against the 0.24 for d_{proj} reported in Appendix E.4. Nor does similarity at the source imply confusion at the receiver. S_{mm} and S_{sum} each contain nine pairs whose source intensities correlate above 0.9, yet their least separable pairs have $\overline{D}$ of 0.941 and 0.909. One of those pairs,

candidates 7 and 429 in S_{mm}, correlates at 0.979, more closely than 99.8% of pool pairs, and has $\overline{D}_{7,429} = 0.990$.

F.4 Conclusion

Within $\mathcal{C}$, selecting for source-intensity diversity under the max–min objective shared with S_{mm} does not reach the binary reference under either dissimilarity. Three limits apply. The control tests two image-space dissimilarities under two objectives, at the single propagation regime of Appendix A.3, and it selects only from the candidate pool $\mathcal{C}$ we construct. That pool was spread by the farthest-point sampling of Equation (6) under d_{proj}, so it covers coefficient space rather than source-intensity space, and a pool built to cover source intensities could offer these criteria candidates that $\mathcal{C}$ lacks. It is also our instantiation, and a different intensity criterion, including that of Jarrett et al. (2024), might order the alphabets differently.