TRINITY: RESIDUAL-AWARE QUADRATIC SEEDING FOR TENSOR-TRAIN BLACK-BOX OPTIMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Trinity connects nonlinear surrogate prediction, bounded quadratic seeding and tensor-train search for discrete black-box optimization. The interface addresses branch cancellation: optimizing the quadratic branch can fail even when the full predictor is exact. For the tested DeepFM model, conditional affine corrections of its one-hidden-layer ReLU residual preserve quadratic couplings without auxiliary bits. A computable residual envelope yields a conditional retained-candidate bound, a near-sharp construction and explicit improvement conditions. Complete seeds are rescored with the full surrogate and initialize a correlated TT; only subsequent true-objective evaluations supply training labels. For the DeepFM instantiation (Trinity-D), across three photonic discretizations (512–2,048 variables), mean independently refined loss falls by 5.0%, 5.9% and 4.8% versus corrected full seeding at matched true calls and quadratic bit proposals. The ridge–DeepFM predictor reduces mean loss by 11.5% relative to Trinity-D on fresh higher-order instances, with a small regression on quadratic instances. Unequal inner work limits attribution, and classical baselines remain stronger in several settings. Recorded costs quantify conditional quantum potential without a measured quantum-speedup claim.

1 INTRODUCTION

Simulator-based design often requires binary material or component choices under a limited evaluation budget. Success depends on both surrogate prediction and surrogate search; increasing model capacity alone does not guarantee either.

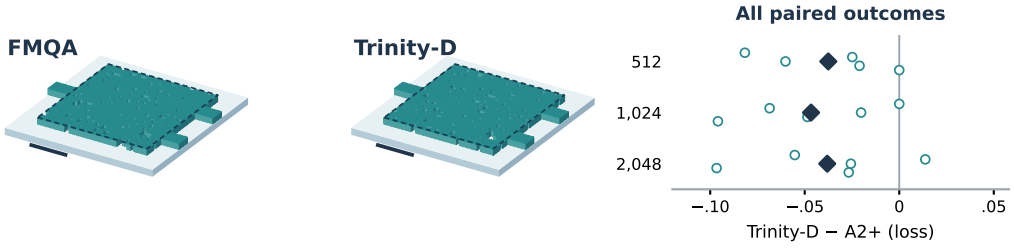

Figure 1: **Saved device layouts and paired evidence.** Left: FMQA and Trinity-D at the prespecified median Trinity-D run (1,024 variables, seed 1002), reconstructed from saved material maps with the same camera and scale. Refined losses are 0.7996 and 0.7310. Teal/pale blue denote $\epsilon_r = 12.25/2.25$; bars are 1 μm and outlines delimit the design region. Thickness is illustrative; physics is two-dimensional. Right: all five paired Trinity-D-minus-A2+ refined-loss differences at each dimension; circles are repetitions and diamonds are means. Negative favors Trinity-D. Mean reductions are 5.0%, 5.9% and 4.8%; these three discretizations share one task family. Table 1 includes stronger external baselines.

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.

FMQA learns a quadratic unconstrained binary optimization (QUBO) model for solver-based search (Kitai et al., 2020). DeepFM adds nonlinear interactions sharing the FM embeddings (Guo et al., 2017); its full prediction is generally nonquadratic.

Optimizing only its quadratic branch can produce wrong seeds even when the full predictor is exact, because the branches can cancel. Rescoring cannot recover designs never proposed.

Trinity asks how a nonlinear surrogate can expose a small-QUBO interface while accounting for its remainder. Fixing passive coordinates leaves at most b active bits. On that subcube, an affine neural approximation changes only the FM's fields and offset and admits a computable error envelope; activation-stable cells make the full-surrogate interface exact. Smaller cells tighten approximation but may exclude better designs.

The contribution is this residual-aware proposal interface and its conditional analysis, integrated with correlated seed-to-TT transfer. PROTES (Batsheva et al., 2023) still searches the full surrogate: warm starting does not remove its quadratic term or imply that PROTES cannot solve QUBOs. The A1–A3 ladder separates unseeded, full-seeded and conditional-seeded search. Complete photonic cohorts support a benefit of the combined seeder across three discretizations of one task, while unequal inner scoring and mixed ablations delimit attribution. Measured runtime identifies the portion a prospective quantum backend could replace.

2 Related work

FMQA couples learned quadratic surrogates to annealing (Kitai et al., 2020); DeepFM combines FM and neural interactions for prediction (Guo et al., 2017). Neither directly establishes that one branch of an additive regressor is a reliable optimization objective. QUBO decomposition and hybrid workflows are established (Bass et al., 2021; D-Wave Systems, 2026). Methods that introduce auxiliary variables for neural or higher-order quadratization (Nakada & Tanaka, 2025) differ from our local affine approximation, which keeps the active bit count unchanged and records its approximation error.

TT representations (Oseledets, 2011) and PROTES (Batsheva et al., 2023) supply established distributional search. The Black–White-Box Optimization Network manuscript already describes an HOFM/Ising/PROTES architecture (Anonymous, 2026). Related pruning and quantization studies seed PROTES with QUBO solutions and then evaluate network performance directly (Orabi et al., 2026b;a). Here PROTES searches a learned g before a separate true-oracle acquisition and refit. Our methodological focus is conditional residual accounting, not the generic QUBO/TT combination. BODi provides a combinatorial Bayesian-optimization comparator (Deshwal et al., 2023).

Affine ReLU relaxations (Wong & Kolter, 2018) and hybrid neural verification (Franco et al., 2022) predate this work. We apply this geometry to bounded seeding, without claiming a new scalar relaxation theorem. Prior DeepFM/QUBO studies also examine quantum execution (Niu et al., 2025); our contribution here is conditional accounting from measured classical costs.

3 Architecture and residual-aware seed generation

Figure 2 shows the nested Trinity loop. A1 uses unseeded DeepFM/PROTES, A2 adds full-QUBO seeds, and A3 uses bounded conditional seeds. The + variants add affine residual correction and observed anchors; *Trinity-D* denotes DeepFM A3+; *Trinity* uses ridge–DeepFM (Section 6). Table 5 gives the full ladder. The *tensor-train creator* (TTC) initializes a TT from complete seeds; PROTES subsequently learns its distribution.

Trinity (A3+): one outer acquisition round
1. Fit $g = q + h$ on all charged observations $\mathcal{D}$; freeze g.
2. Repeat controller passes: partition active bits; sample passive contexts; build corrected local QUBOs; solve, lift and archive complete vectors. Rescore with g, apply genetics, retain elites and update context probabilities.
3. Remove observed designs; rank/diversify seeds and construct P_0 (default TT if empty).
4. Repeat PROTES sampling, full-g scoring and TT updates within its budget.
5. Select the best unseen samples; fill any shortfall with random unseen designs. Query f and append *all* new labels to $\mathcal{D}$. Repeat until the true-call budget is exhausted.

3.1 Problem, predictor and exact restrictions

We minimize a deterministic $f : \{0,1\}^d \to \mathbb{R}$ using at most N actual objective calls, including initialization. A dataset $\mathcal{D}$ contains charged observations. The two nested input–output oracles are the true objective f and the fitted model g. PROTES queries full g as a black box; the seeder accesses its structure. Every outer round follows the boxed procedure. We distinguish true calls N_f, full-model scores N_g, and QUBO solves N_Q; only calls to f provide new training labels. The analysis concerns unconstrained binary decisions; feasibility repair or noisy objectives would require additional assumptions.

The architecture accepts $g = q + h$ with accessible quadratic q. Theorem 1 is model-independent, but the computable ReLU correction is specific. Appendix A.1 gives an untested HOFM extension (Blondel et al., 2016). We use DeepFM (Trinity-D) and a ridge–DeepFM predictor (Trinity). Both expose the prediction in objective units

$$g(x) = q(x) + h(x), \qquad q(x) = c + w^\top x + \sum_{i<j} \langle v_i, v_j \rangle x_i x_j, \tag{1}$$

where $v_i \in \mathbb{R}^k$ is the extracted quadratic factor. A one-hidden-layer ReLU network receives flattened gated training embeddings $(x_i e_i)_{i=1}^d$, shared with the unscaled FM. There is no output sigmoid. At evaluation, dropout is disabled and the neural branch can be collapsed exactly to

$$h(x) = \gamma + \sum_{\ell=1}^{H} a_\ell \,\mathrm{ReLU}(c_\ell + C_\ell x). \tag{2}$$

For Trinity, a training-only centered ridge model ($\lambda = 10$) supplies a linear backbone. DeepFM fits its residual; validation chooses this sum or ridge alone. The backbone is folded into c, w in (1). For the DeepFM component, target standardization is undone in both branches: if $y = (t - \mu)/\sigma$ for its training target t, then $q = \mu + \sigma q_{\mathrm{std}}$ and $h = \sigma h_{\mathrm{std}}$. Extraction gives $v_i = \sqrt{\sigma} e_i$; the neural preactivations use the unscaled e_i.

For active coordinates A, passive coordinates P, and context z_P, write $\mathcal{C}(A, z) = \{x : x_P = z_P\}$ and let $L_{A,z}(u)$ lift a local assignment. Directly partitioning the FM terms gives

$$q(L_{A,z}(u)) = c_z + \sum_{i \in A} \left(w_i + \left\langle v_i, \sum_{j \in P} v_j z_j \right\rangle \right) u_i + \sum_{i<j;\ i,j \in A} \langle v_i, v_j \rangle u_i u_j. \tag{3}$$

Passive cross-interactions become fields, not discarded interactions. Candidates from different contexts are lifted and scored as complete vectors; local energies without their offsets are not comparable across contexts. A fair full-QUBO classical baseline also exploits the factorization, using $O(dk)$ storage and $O(k)$ work per proposed bit flip.

3.2 Correcting the neural branch without extra bits

On $\mathcal{C}(A, z)$, each preactivation t_ℓ has the exact box interval

$$L_\ell = c_\ell + C_{\ell P} z_P + \sum_{i \in A} \min(0, C_{\ell i}), \qquad U_\ell = c_\ell + C_{\ell P} z_P + \sum_{i \in A} \max(0, C_{\ell i}). \tag{4}$$

Use the exact affine function for an activation-stable unit. If $L_\ell < 0 < U_\ell$, use the secant $s_\ell(t) = U_\ell(t - L_\ell)/(U_\ell - L_\ell)$ and define $E_\ell = -L_\ell U_\ell/(U_\ell - L_\ell)$. Then $0 \le s_\ell(t) - \mathrm{ReLU}(t) \le E_\ell$.

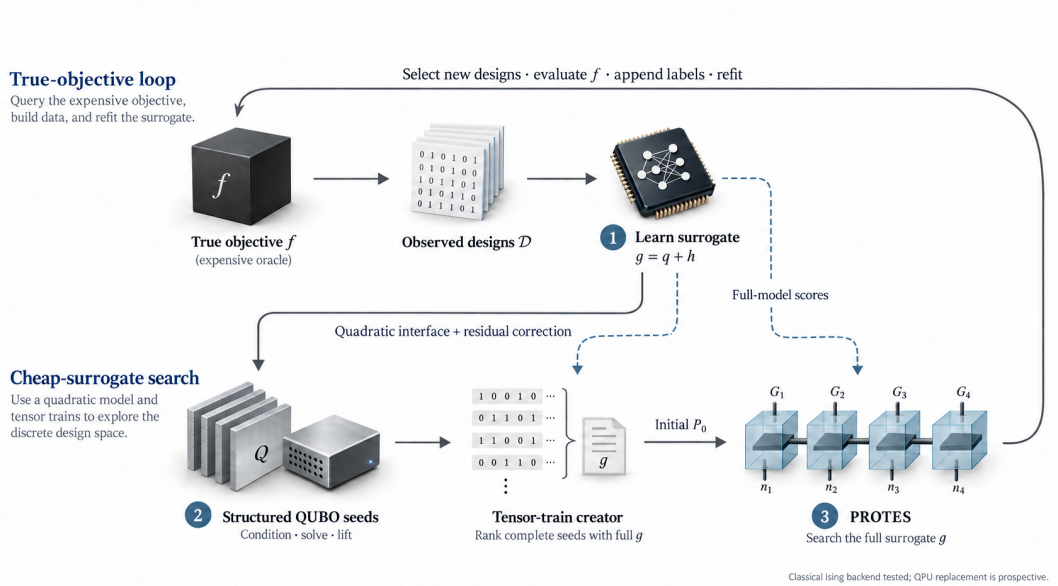

Figure 2: Trinity's nested optimization loop. The true oracle supplies labels to the AI surrogate; its quadratic interface generates complete seeds that are ranked with full g before the tensor-train creator initializes P_0. PROTES repeatedly samples, scores the frozen full $g = q + h$, and updates its TT. Distinct unseen candidates return to the true oracle, and all new labels enter the next fit. Context-probability updates belong to the seeder, not PROTES. Experiments use DeepFM and a classical Ising solver; a QPU backend is prospective. Figure 4 expands seeding; Table 5 compares initializations.

Replacing the ReLUs in (2) gives an affine $\widetilde{h}_{A,z}$ and the corrected QUBO

$$\psi_{A,z}(u) = q(L_{A,z}(u)) + \widetilde{h}_{A,z}(u). \tag{5}$$

The quadratic couplings are unchanged. The correction adds no auxiliary bits and applies to this network architecture, not an arbitrary deep network. For $r = h - \widetilde{h}$, the signed interval is

$$r_- = -\sum_{a_\ell > 0} a_\ell E_\ell \le r \le r_+ = \sum_{a_\ell < 0} |a_\ell| E_\ell, \qquad B(A,z) = r_+ - r_- = \sum_{\ell} |a_\ell| E_\ell. \tag{6}$$

Mixed output signs prevent treating the affine approximation itself as a universal upper bound. B bounds represented-model residual oscillation; it is not an estimate of prediction error against f.

3.3 Contexts, archive, genetics and TT transfer

A3 starts with product-controller probabilities $p_i = 1/2$. Each controller pass draws a new random permutation and partitions coordinates into blocks of at most b. Passive contexts are sampled from the controller. A3+ mixes this distribution with perturbations of the eight best observed designs: with probability $1/2$, choose an anchor uniformly and flip each bit with probability $1/d$. Full-seeding A2+ receives the same observed anchors for its annealing starts. An improvement from anchor access alone is therefore not credited to decomposition.

For each subproblem, retain distinct local solver candidates, lift them with their own contexts, and insert complete candidates into a bounded archive *before* genetic parent selection. Rank them using g, preserve the archive's best full-score point, and apply one generation of crossover/mutation to complete vectors. Update the controller using quality/diversity elites,

$$p_i \leftarrow \mathrm{clip}_{[\delta, 1-\delta]} \left((1-\alpha) p_i + \alpha \, \overline{x}_i^{\mathrm{elite}} \right). \tag{7}$$

Archive truncation and the local retention rule limit which witnesses are available to the theory. Genetics uses surrogate scores only.

The final distinct seed set $S = \{s_1, \ldots, s_m\}$ excludes already evaluated designs. We initialize a correlated distribution

$$P_0(x) = \omega 2^{-d} + \frac{1-\omega}{m} \sum_{j=1}^{m} \prod_{i=1}^{d} \left[(1-\rho)\mathbf{1}\{x_i = s_{ji}\} + \rho \mathbf{1}\{x_i \neq s_{ji}\} \right], \tag{8}$$

represented exactly by nonnegative TT cores of rank at most $m+1$. We use $m \leq 7$, rank eight, $\omega = 0.1$, and $\rho = 1/d$. This is a mixture of complete-seed neighborhoods, not the product of their one-bit marginals. PROTES then updates the TT while scoring the same fixed full surrogate. A batch of distinct new designs is selected from evaluated TT candidates; the primary method does not silently send archive elites directly to the oracle. The product controller and final TT serve different roles; Appendix B.2 details their placement.

4 What conditional quadratic seeds can guarantee

4.1 Candidate quality, and why branch accuracy matters

Theorem 1 (Retained conditional witness)**.** *For each sampled subcube $\mathcal{C}_j$, suppose a retained candidate y_j satisfies $\psi_j(y_j) \leq \min_{\mathcal{C}_j} \psi_j + \eta_j$, with $\eta_j \geq 0$, and $\operatorname{osc}_{\mathcal{C}_j}(g - \psi_j) \leq B_j$. Let the archive contain these candidates and let $\bar{x}$ minimize g over the retained archive. If $|f - g| \leq \epsilon$ on all these subcubes and the retained archive, including genetic additions, then*

$$f(\bar{x}) \leq \min_j \left[\min_{x \in \mathcal{C}_j} f(x) + \eta_j + B_j \right] + 2\epsilon. \tag{9}$$

Proof. For a true conditional minimizer x_j^*, archive retention and full-score selection imply $g(\bar{x}) \leq g(y_j) \leq \psi_j(x_j^*) + \eta_j + (g - \psi_j)(y_j) \leq g(x_j^*) + \eta_j + B_j$. Apply the prediction-error bound at x_j^* and $\bar{x}$, then minimize over j. □

The bound is conditional on a useful sampled subcube and a retained solver witness. It does not certify a global optimum or a later adaptive TT sample. Validation R^2 does not establish its uniform ϵ assumption. For raw seeding, replace B_j by $\operatorname{osc}_{\mathcal{C}_j} h$. Appendix A.3 gives a four-bit, one-ReLU construction approaching all three additive terms simultaneously.

An exact full predictor can still give wrong raw seeds. Let $f(x) = x_1$, $q(x) = -Mx_1$, and $h(x) = (M+1)\,\mathrm{ReLU}(x_1)$, with $M > 0$. Then $g = f$ but minimizing q chooses the wrong bit. The neural unit is stable on $[0, 1]$, so the correction repairs this example exactly. Prediction accuracy alone does not identify the quadratic branch as a useful search objective.

4.2 Smaller cells tighten the envelope, not necessarily the objective

Proposition 1 (Certificate scale)**.** *For the fixed represented network, B cannot increase when a subcube is restricted by fixing additional coordinates consistently. Moreover, with $W_A = \frac{1}{4}\sum_\ell |a_\ell| \sum_{i \in A} |C_{\ell i}|$ and $W = W_{[d]}$,*

$$B(A, z) \leq W_A, \qquad \mathbb{E}_A W_A = \frac{b}{d} W \tag{10}$$

when A is a uniform b-subset, even if the context depends on A.

The proof uses $E_\ell \leq (U_\ell - L_\ell)/4$ and each coordinate's inclusion probability b/d. On a crossing interval write $u = -L, v = U$; $uv/(u+v)$ increases in both nonnegative arguments, establishing nested-cell monotonicity. Activation-stable cells have $B = 0$. The factor b/d applies to W, not necessarily to the actual full-cube certificate. Different independently sampled contexts need not be nested.

Figure 3 illustrates strict separation from global correction. More generally, smaller blocks also reduce the range of coordinated changes and can miss a better basin. A useful improvement statement must include this context tradeoff. Fix an observed anchor z^* and suppose a random b-subset admits conditional true improvement at least Δ with probability at least κ. For $0 < \xi < \kappa$, let

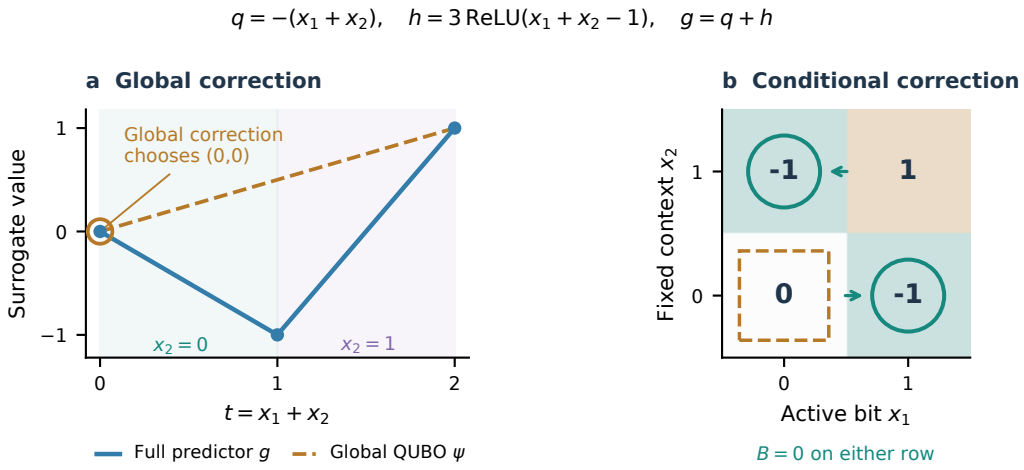

Figure 3: Exact mathematical example, not experimental performance. For $g(x) = -(x_1 + x_2) + 3\,\mathrm{ReLU}(x_1 + x_2 - 1)$, global secant correction chooses $(0, 0)$, while either one-bit conditional problem has an exact neural approximation and returns a one-hot minimizer. The binary function itself is quadratic; the example separates two approximations, not conditional search from an exact solver of the true full objective.

$\tau = bW/(d\xi)$. Suppose the context mixture assigns this anchor weight β/M and the solver reaches accuracy η with conditional probability at least s on every relevant cell. The union and Markov bounds give a retained archive improvement

$$f(\bar{x}) \le f(z^*) - \Delta + \eta + \tau + 2\epsilon \tag{11}$$

with single-trial probability at least $p = (\kappa - \xi)(\beta/M)(1 - \rho_c)^{d-b}s$. Strict improvement requires $\Delta > \eta + \tau + 2\epsilon$. These quantities are explanatory assumptions, not inferred physical guarantees. Appendix A.4 distinguishes history-wise success probabilities from marginal coverage within a random partition.

4.3 Recovering seeds and bounding solver error

Each supplied seed in (8) has mass at least $p_{\mathrm{seed}} = (1 - \omega)(1 - \rho)^d/m$. For K independent draws from this fixed initial distribution, recovery probability is at least $1-(1-p_{\mathrm{seed}})^K$. With $\rho = c/d$, the bound avoids exponential decay in dimension for fixed c and m. It does not assert that the supplied seed is good, unseen, or retained by subsequent adaptive updates. In contrast, the mixture on 0^d and 1^d has TT rank two, while its independent uniform marginals assign their union mass only 2^{1-d} (Appendix A.5).

A matrix-free quadratic lower bound can diagnose solver error after a solve (Appendix A.6); it is not part of the frozen search. The guarantee chain is deliberately limited: correction controls error relative to g on a fixed cell; Theorem 1 controls a retained archive witness under uniform error relative to f; the TT bound controls initial sampling only if that witness survives eligibility and seed selection. No bound here covers the final adaptive PROTES output. Floating-point diagnostics lack outward rounding and are not machine-certified bounds.

5 Experimental design

We distinguish development data, frozen final protocols, and additional physical validation. The original frozen inventory contains 870 planned runs across ten studies; subsequent optical follow-ups (70 runs) and a 20,000-variable, 100-label study (36 runs) are identified separately in Appendix B.1. We report completed, audited endpoints; incomplete cohorts are retained in the inventory and excluded from aggregate comparisons. All methods in a paired condition share initial binary designs

and objective values. Source snapshots, configuration hashes, unique-query journals, checkpoints, failures and runtime breakdowns are retained. No pilot is retrospectively relabeled as a final test.

Objectives. The controlled family uses a hidden bit vector x^* and $z = x \oplus x^*$. Its loss is $0.1\,\overline{z} + 0.9[(1-\nu)D_2(z) + \nu D_4(z)]$, where D_2 averages pair disagreements and D_4 averages four-bit OR terms. All terms vanish at x^*, and the positive linear term makes it the unique optimum. We use $\nu = 0, 0.5$ at $d = 512, 1024, 2048$, and $\nu = 0.5$ at $d = 5000, 10000$. The higher-order condition contains degrees up to four. This is a planted, simultaneously satisfiable family, not a frustrated spin glass or a real application.

The principal application is a two-dimensional wavelength demultiplexer simulated with Ceviche and Ceviche Challenges (Hughes et al., 2019; Google Research, 2026). Binary material maps have 512, 1024, or 2048 genuine design bits. The objective penalizes desired-port transmission error and undesired-port transmission at 1270 and 1290 nm. Optimization uses a 20 nm grid; every terminal design and common initial best is reevaluated at 10 nm with fixed 200 nm monitor offsets. The finer-grid response never feeds back into selection. Linear-system residuals and complex scattering values support independent audits.

A separate low-dimensional pilot chooses the material of each of 32 planar 75 nm layers. Coherent transfer-matrix simulation uses tabulated complex TiO_2 optical constants and the fused-silica Sellmeier formula (Siefke et al., 2016; Malitson, 1965). The target is low reflectance over 450–550 nm and high reflectance over 650–750 nm, with 11 wavelengths per band. These are data-grounded simulations, not fabricated-device measurements. In particular, transferring measured film constants to the modeled layer thickness is an assumption.

Methods and budgets. The primary ladder is FMQA-style A0, an FM/PROTES bridge, unseeded DeepFM/PROTES A1, full-QUBO-seeded A2, and bounded-seeded A3. Anchored FM and corrected/anchored A2+ and A3+ control anchor access and correction, but do not match total inner scoring. Random search, a cross-entropy method (CEM), genetic search, direct PROTES and BODi provide controls. BODi uses pinned authors' dictionary and acquisition-search code with verified algebraically equivalent binary encoding and marginal-EI evaluation. It is sequential; our main surrogate methods query batches of 32. Both pay for the same initial observations and total true calls. This compares sample efficiency, not matched batch latency or cross-host wall time.

Final budgets are 256 calls, with 128 shared initial designs except 64 in the film pilot. There are five paired instances/seeds per high-dimensional condition and ten for film. The high-dimensional setting leaves four adaptive rounds after initialization. Default active blocks have 64 bits (eight for film). Appendix B.1 gives training, solver, TT and ablation settings. BODi is included in the 512–2048 studies and film, but not the separate 5000–10000 scaling study. We report all planned failures and descriptive paired uncertainty; five seeds do not justify broad significance claims or pooling dependent instances across dimensions.

6 Results and Diagnostic Evidence

6.1 Where Trinity improves surrogate search

Trinity-D's conditional interface improves mean independently refined photonic loss over corrected full seeding at matched true calls and quadratic bit proposals (A2+) by 5.0%, 5.9%, and 4.8% at 512, 1,024, and 2,048 variables (Figure 1; Table 1). At 1,024 variables, mean loss falls from 0.78572 to 0.73906, with four wins and one tie. The reduction relative to FMQA is 4.4%, also four wins and one tie. Across complete comparable cohorts, Trinity-D has lower mean loss in **12/12** conditions versus direct PROTES, **12/12** versus A1, **6/10** versus FMQA, and **8/12** versus A2+. These are descriptive condition counts, not significance or overall first-place counts. Against GA, CEM and BODi the corresponding counts are 5/12, 4/12 and 0/7.

Paired A3+ minus A2+ refined-loss differences are -0.03754, -0.04666, and -0.03809 across the three sizes, with descriptive 95% bootstrap intervals $[-0.06523, -0.01338]$, $[-0.07552, -0.01775]$, and $[-0.07181, -0.00814]$. With only five pairs, even five non-tied wins give exact two-sided sign-test $p = 0.0625$; four wins and a tie give $p = 0.125$. Intervals and counts are therefore exploratory. On the controlled $d = 2{,}048$, $\nu = 0$ instance family, Trinity-D attains the lowest mean among

Table 1: Mean terminal loss at 256 true calls; lower is better within a row. Bold marks the lowest unrounded mean among displayed complete cohorts. Five paired repetitions per condition (ten for film); WDM uses independent 10 nm reevaluation. PROTES queries f directly; A1 searches DeepFM. T-D denotes DeepFM Trinity-D (A3+). Panels group conditions by complete comparators: the middle excludes BODi; the bottom excludes methods not planned for scaling. Dispersion and every complete method appear in Appendix C.

Condition	FMQA	FM+anc.	PROTES	Rand.	CEM	GA	BODi	A1	A2+	T-D
Film (32)	0.0913	0.0902	0.0826	0.0982	0.0707	0.0560	**0.0136**	0.0844	0.0719	0.0715
512, $\nu = 0$	0.4539	0.4526	0.4646	0.4655	0.4622	0.4558	**0.4227**	0.4644	0.4592	0.4630
512, $\nu = 0.5$	0.6055	0.6073	0.6631	0.6627	0.6399	0.6494	**0.5926**	0.6616	0.6088	0.6251
1,024, $\nu = 0$	0.4760	0.4704	0.4760	0.4753	0.4767	0.4715	**0.4504**	0.4773	0.4773	0.4672
1,024, $\nu = 0.5$	0.6143	**0.6094**	0.6738	0.6746	0.6542	0.6628	0.6197	0.6719	0.6305	0.6575
WDM 512	0.7126	0.7301	0.7420	0.7543	0.7044	0.6994	**0.5527**	0.7500	0.7463	0.7087
WDM 1,024	0.7729	0.7640	0.7857	0.7857	0.7582	0.7342	**0.6594**	0.7857	0.7857	0.7391

Condition	FMQA	FM+anc.	PROTES	Rand.	CEM	GA	A1	A2+	T-D
2,048, $\nu = 0$	0.4817	0.4797	0.4829	0.4845	0.4832	0.4807	0.4849	0.4828	**0.4795**
2,048, $\nu = 0.5$	**0.6477**	0.6505	0.6797	0.6814	0.6712	0.6755	0.6806	0.6636	0.6733
WDM 2,048	0.7993	0.7458	0.7722	0.7991	**0.7330**	0.7335	0.7991	0.7991	0.7610

Condition	FM+anc.	PROTES	CEM	GA	A1	A2+	T-D
5,000	**0.6819**	0.6874	0.6826	0.6834	0.6885	0.6860	0.6853
10,000	0.6891	0.6901	**0.6871**	0.6880	0.6894	0.6898	0.6892

Table 2: Frozen component controls at 512 optical variables: five paired seeds, 256 true calls, mean $\pm$ sample SD after 10 nm reevaluation. Δ is variant minus the stated reference; negative favors the variant. Bold marks the lowest mean among these configurations. Block-size changes also change solver work.

Setting	Reference	10 nm loss	Δ vs. reference
A3+ (reference)	—	0.70872 ± 0.04061	—
No neural correction	A3+	0.73580 ± 0.05562	$+0.02709$
No observed anchors	A3+	0.75961 ± 0.03737	$+0.05089$
No genetics	A3+	$\mathbf{0.69364 \pm 0.05628}$	-0.01507
No TT updates	A3+	0.74621 ± 0.04492	$+0.03749$
Archive capacity 32	A3+	0.71474 ± 0.03363	$+0.00603$
Block size 16	A3+	0.70250 ± 0.08231	-0.00622
Block size 512	A3+	0.71808 ± 0.06511	$+0.00936$
Raw A3 (reference)	—	0.75450 ± 0.02885	—
Fixed controller	Raw A3	0.73894 ± 0.02683	-0.01556

twelve completed methods: 0.47950 versus 0.47966 for anchored FMQA. The 0.00016 margin is not a demonstrated significant advantage. BODi is excluded there because its planned cohort is incomplete. All method comparisons, including external-baseline wins, remain in the table and Appendix C.

6.2 What the component controls establish

All 40 optical ablation runs and refinements completed (Table 2). Removing observed anchors or TT updates increases mean refined loss by 0.05089 or 0.03749 relative to A3+. Removing genetics instead lowers the mean by 0.01507; its descriptive paired interval spans zero. The fixed-controller comparison uses raw A3, matching its actual implementation. These mixed effects do not justify claiming every component is necessary. Settings were not retuned after these outcomes; Appendix C.7.1 supplies paired differences and both grids.

Table 3: Fresh 512-bit instances: five paired seeds, 256 true calls, 128 shared starts. Mean loss; lower is better, best in bold. T-D is the DeepFM control; A2 shares Trinity's ridge–DeepFM predictor. Validation covers the same synthetic family.

ν	FM+anc.	PROTES	CEM	GA	A1	T-D	A2 (ridge)	Trinity
0	**0.4522**	0.4686	0.4630	0.4559	0.4635	0.4621	0.4700	0.4658
0.5	0.5945	0.6634	0.6426	0.6446	0.6631	0.6391	**0.5623**	0.5653

6.3 RIDGE–DEEPFM TRINITY ON FRESH INSTANCES

The fixed predictor was tested in 20 development and 80 fresh-instance runs (Table 3; Appendix C.10).

On fresh higher-order instances, Trinity reduces mean loss from 0.63910 to 0.56531 (11.5%), winning all five pairs against Trinity-D. A2 (ridge) reaches 0.56234, so decomposition does not uniquely explain the gain. Quadratic loss increases by 0.00371. This validates fresh instances of one selected synthetic family; neither applications nor BODi were replicated. The appendix retains all development outcomes.

DeepFM A3+ uses $17.7\times$ A2+'s full-model scores at 10,000 variables (Table 18); neither study matches total inner work. Optical, ablation and cost measurements use Trinity-D, with stronger BODi/GA optical means. Appendix C.8 reports physical metrics; incomplete cohorts are excluded from comparisons.

7 CLASSICAL COST AND A POSSIBLE QUANTUM BACKEND

For R rounds, T passes and C contexts, seeding submits $RTC\lceil d/b \rceil$ QUBOs. All remaining work stays on the host. Measured solver fraction θ, assumed acceleration s and added overhead H_Q normalized by original runtime give

$$S = \left[(1-\theta) + \theta/s + H_Q\right]^{-1}. \tag{12}$$

A hypothetical quality-matched, all-in 1 ms solve gives 1.44–$1.62\times$ at 5,000–10,000 bits; documented programming alone typically takes 10–20 ms (D-Wave Quantum, 2026b). Zero-cost replacement caps speedup at 1.74–1.82 (5,000 bits), 1.57–1.60 (10,000), and 1.018–1.021 (optical 1,024). Table 19 and Appendix E detail measured costs and hardware overheads. These conditional scenarios establish no quantum advantage.

8 LIMITATIONS AND FUTURE WORK

Guarantees assume useful contexts, accurate solves and uniform surrogate error bounds. Missing controls include matched inner work, anchor-only search and same-surrogate GA/CEM. The scaling family is planted; the film pilot is low-dimensional. Photonic physics is 2D and nondispersive, without fabrication validation.

Future work targets compute-matched controls, optical replication, improved scarce-label prediction, HOFM and stable TT updates. Generalization requires new task families, fabrication-aware 3D physics and quality-matched QPU experiments including host costs.

9 CONCLUSION

Trinity combines residual-aware quadratic seeding and nonlinear-surrogate TT search. Conditional theory explains retained-seed quality. Trinity-D improves on full seeding across three optical discretizations, while ridge–DeepFM Trinity improves fresh higher-order results. Unequal inner work and stronger baselines limit attribution. Recorded costs quantify a potential backend acceleration; broader superiority and quantum advantage require matched evaluation.

USE OF GENERATIVE AI

Generative AI assisted research planning and hypothesis refinement, mathematical claims and proof drafting, synthetic benchmark definition, experimental design, code implementation and debugging, literature retrieval, experiment orchestration, analysis and interpretation of results, manuscript drafting, and figure preparation. Conceptual architecture illustrations are AI-generated and checked against the stated algorithm; their objects do not depict measured hardware or experimental observations. Empirical plots derive from stored observations, and mathematical illustrations are explicitly labeled.

REFERENCES

Anonymous. The black–white-box optimization network, 2026. URL https://openreview.net/forum?id=JrZMFC6Jgo. Public manuscript bearing an ICLR 2026 review header; publication status not asserted.

Gideon Bass, Maxwell Henderson, Joshua Heath, and Joseph Dulny, III. Optimizing the optimizer: decomposition techniques for quantum annealing. *Quantum Machine Intelligence*, 3:10, 2021. doi: 10.1007/s42484-021-00039-9.

Anastasiia Batsheva, Andrei Chertkov, Gleb Ryzhakov, and Ivan Oseledets. PROTES: Probabilistic optimization with tensor sampling. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-0037. URL https://arxiv.org/abs/2301.12162.

Mathieu Blondel, Akinori Fujino, Naonori Ueda, and Masakazu Ishihata. Higher-order factorization machines. In *Advances in Neural Information Processing Systems*, volume 29, 2016. URL https://papers.nips.cc/paper_files/paper/2016/hash/158fc2ddd52ec2cf54d3c161f2dd6517-Abstract.html.

Steven J. Byrnes. Multilayer optical calculations. arXiv:1603.02720, 2016. URL https://arxiv.org/abs/1603.02720. Version 5, 2020; formulas underlying the tmm software.

D-Wave Quantum. Parallelembeddingcomposite, 2026a. URL https://docs.dwavequantum.com/en/latest/ocean/api_ref_system/composites.html. Ocean SDK documentation; accessed September 26, 2026.

D-Wave Quantum. Operation and timing, 2026b. URL https://docs.dwavequantum.com/en/latest/quantum_research/operation_timing.html. Quantum Computing Products documentation; accessed September 26, 2026.

D-Wave Systems. Decomposing problems. Ocean documentation, 2026. URL https://docs.dwavequantum.com/en/latest/quantum_research/decomposing.html. Accessed September 26, 2026.

Aryan Deshwal, Sebastian Ament, Maximilian Balandat, Eytan Bakshy, Janardhan Rao Doppa, and David Eriksson. Bayesian optimization over high-dimensional combinatorial spaces via dictionary-based embeddings. In *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics*, volume 206, pp. 7021–7039. PMLR, 2023. URL https://proceedings.mlr.press/v206/deshwal23a.html.

Nicola Franco, Tom Wollschläger, Nicholas Gao, Jeanette Miriam Lorenz, and Stephan Günnemann. Quantum robustness verification: A hybrid quantum-classical neural network certification algorithm, 2022. URL https://arxiv.org/abs/2205.00900.

John Gardiner and Javier Lopez-Piqueres. Tensor network estimation of distribution algorithms. *Neurocomputing*, 680:133255, 2026. doi: 10.1016/j.neucom.2026.133255. URL https://arxiv.org/abs/2412.19780.

Google Research. Ceviche challenges, 2026. URL https://github.com/google/ceviche-challenges. Software, version 1.0.2; accessed September 26, 2026.

Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. DeepFM: A factorization-machine based neural network for CTR prediction. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 1725–1731, 2017. doi: 10.24963/ijcai.2017/239. URL https://www.ijcai.org/Proceedings/2017/239.

Tyler W. Hughes, Ian A. D. Williamson, Momchil Minkov, and Shanhui Fan. Forward-mode differentiation of Maxwell's equations. *ACS Photonics*, 6(11):3010–3016, 2019. doi: 10.1021/acsphotonics.9b01238. URL https://github.com/fancompute/ceviche.

Koki Kitai, Jiang Guo, Shenghong Ju, Shu Tanaka, Koji Tsuda, Junichiro Shiomi, and Ryo Tamura. Designing metamaterials with quantum annealing and factorization machines. *Physical Review Research*, 2:013319, 2020. doi: 10.1103/PhysRevResearch.2.013319.

I. H. Malitson. Interspecimen comparison of the refractive index of fused silica. *Journal of the Optical Society of America*, 55:1205–1209, 1965. doi: 10.1364/JOSA.55.001205.

Hyakka Nakada and Shu Tanaka. Systematic and efficient construction of quadratic unconstrained binary optimization forms for high-order and dense interactions. *Journal of the Physical Society of Japan*, 94(9):094801, 2025. doi: 10.7566/JPSJ.94.094801. URL https://arxiv.org/abs/2506.08448.

Jiayang Niu, Qihan Zou, Jie Li, Ke Deng, Mark Sanderson, and Yongli Ren. Estimating quantum execution requirements for feature selection in recommender systems using extreme value theory, 2025. URL https://arxiv.org/abs/2507.03229v2. Accepted to the RecSys 2025 short-paper track.

Osama Orabi, Artur Zagitov, Hadi Salloum, Viktor A. Lobachev, and Yaroslav Kholodov. Task-aware qubo allocation for mixed-precision quantization, 2026a. URL https://arxiv.org/abs/2609.22238v2. Version 2, September 23, 2026.

Osama Orabi, Artur Zagitov, Hadi Salloum, Viktor A. Lobachev, Kasymkhan Khubiev, and Yaroslav Kholodov. Neural network pruning via qubo optimization, 2026b. URL https://arxiv.org/abs/2604.05856.

Ivan V. Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011. doi: 10.1137/090752286.

Thomas Siefke, S. Kroker, K. Pfeiffer, O. Puffky, K. Dietrich, D. Franta, I. Ohlídal, A. Szeghalmi, E.-B. Kley, and A. Tünnermann. Materials pushing the application limits of wire grid polarizers further into the deep ultraviolet spectral range. *Advanced Optical Materials*, 4:1780–1786, 2016. doi: 10.1002/adom.201600250.

Eric Wong and Zico Kolter. Provable defenses against adversarial examples via the convex outer adversarial polytope. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pp. 5286–5295. PMLR, 2018. URL https://proceedings.mlr.press/v80/wong18a.html.

A MATHEMATICAL DETAILS

A.1 THE SURROGATE INTERFACE BEYOND DEEPFM

Trinity's architectural requirement is a cheap predictor g with an accessible quadratic proposal model, a full-score evaluator and a seed-to-TT interface. It is not a claim that every fitted black box exposes a useful QUBO. Theorem 1 applies to any proposal ψ with a valid residual-oscillation bound; an arbitrary predictor need not admit an efficiently computable bound.

For example, a finite-order HOFM represents a multilinear polynomial (Blondel et al., 2016). Write its represented function as $g(x) = \sum_{S\subseteq[d]} c_S \prod_{i\in S} x_i$, with zero coefficients above its order. On a fixed passive context,

$$g(L_{A,z}(u)) = \sum_{T\subseteq A} a_T(z) \prod_{i\in T} u_i, \qquad a_T(z) = \sum_{S:S\cap A=T} c_S \prod_{j\in S\cap P} z_j. \tag{13}$$

The terms with $|T| \le 2$ give a conditional QUBO $\psi_{A,z}$; the remaining terms give $r_{A,z}$. Each binary monomial lies in $[0, 1]$, so

$$\operatorname{osc}_{\mathcal{C}(A,z)} r_{A,z} \le \sum_{T\subseteq A:\, |T|\ge 3} |a_T(z)|. \tag{14}$$

This algebraic bound follows by summing individual term ranges. Fixing passive bits can convert higher-order interactions into quadratic or linear terms, which should be retained. For at most two active bits the full represented function is exactly quadratic. Unlike the specific affine ReLU correction, this contraction can change quadratic couplings. Explicit coefficient expansion may be prohibitive; a practical HOFM adapter would need to exploit its factorization. This is an untested extension, not a reported implementation or empirical result. In every case, PROTES continues to score the full g, and only evaluations of f augment the training data.

A.2 EXACT IDENTITIES AND CONDITIONAL PROOF DETAILS

For binary x, expanding the squared embedding sum gives

$$q(x) = c + w^\top x + \tfrac{1}{2}\left(\|V^\top x\|^2 - \sum_i \|v_i\|^2 x_i\right). \tag{15}$$

With $s = V^\top x$, the energy change from flipping bit i is $(1 - 2x_i)(w_i + v_i^\top s - \|v_i\|^2 x_i)$; update s by $(1 - 2x_i)v_i$. This identity explains the matrix-free classical solver and the need to give the complete-QUBO comparator the same structure-aware implementation.

For a local binary energy $c + \sum_i a_i u_i + \sum_{i<j} B_{ij} u_i u_j$, substitute $u_i = (s_i + 1)/2$. The exact Ising coefficients are

$$J_{ij} = B_{ij}/4, \quad H_i = a_i/2 + \tfrac{1}{4}\sum_{j\ne i} B_{ij}, \quad C = c + \tfrac{1}{2}\sum_i a_i + \tfrac{1}{4}\sum_{i<j} B_{ij}. \tag{16}$$

Here B_{ij} denotes an unordered-pair coefficient with symmetric lookup. A dense $u^\top Q u$ convention has a different off-diagonal factor. Complete enumeration checks the implemented conventions, offsets and lifting.

For the neural correction, each collapsed coefficient is $C_{\ell i} = W_{\ell i} e_i$, where e_i is the unscaled training embedding and $W_{\ell i}$ its hidden-layer weight block. In objective units $a_\ell = \sigma w_{\mathrm{out},\ell}$ and $\gamma = \sigma b_{\mathrm{out}}$; the hidden preactivation remains in its network units. On a crossing interval the secant error increases up to zero and decreases thereafter, attaining $-LU/(U - L)$ at zero. Since $-LU \le (U - L)^2/4$, its bound is no larger than one quarter of the interval width. If $L = U$ a stable branch avoids division by zero. Multiplying the scalar errors by signed output weights yields (6); different units need not attain their individual maxima simultaneously, so the summed bound can be loose.

A.2.1 A STRICT SEPARATION FROM GLOBAL CORRECTION

Let $q(x) = -a(x_1 + x_2)$ and $h(x) = K\,\mathrm{ReLU}(x_1 + x_2 - 1)$, where $K > 2a > 0$. At the four binary states, g equals $0, -a, -a, K-2a$. The global neural secant is $K(x_1+x_2)/2$, so the corrected

global quadratic has a unique minimizer at $(0, 0)$. If either coordinate is fixed, the remaining unit's interval is $[-1, 0]$ or $[0, 1]$. The correction is exact on either interval and returns a one-hot minimizer. Thus the corresponding local construction can improve the full-correction seed despite exact QUBO solves and zero prediction error. This is an explanatory approximation example: on binary inputs $\mathrm{ReLU}(x_1 + x_2 - 1) = x_1 x_2$, so an exact solver of the true full quadratic would also find the optimum.

A.3 Near-sharpness of the candidate bound

Choose $\epsilon > 0$, $B > 0$, $\eta \geq 0$, $0 < \delta_B < B$ and $0 < \delta_e < 2\epsilon$. Set $a = B - \delta_B$, $K = 2B$, and use four active bits (u, v, e, s). Define

$$f = a - a(u + v) + K\,\mathrm{ReLU}(u + v - 1) + (2\epsilon - \delta_e)e + \eta s, \tag{17}$$

$$g = a - a(u + v) + K\,\mathrm{ReLU}(u + v - 1) + \epsilon - \delta_e e + \eta s. \tag{18}$$

Then $g - f = \epsilon - 2\epsilon e$, establishing uniform error exactly ϵ. The true minimum is zero at a one-hot (u, v) with $e = s = 0$. The secant of the unit over $[-1, 1]$ is $(u + v)/2$, so

$$\psi = a + \epsilon + (B - a)(u + v) - \delta_e e + \eta s. \tag{19}$$

Its minimum occurs at $(0, 0, 1, 0)$. The returned candidate $y = (0, 0, 1, 1)$ is exactly η-suboptimal for ψ, and $g - \psi$ attains both 0 and $-B$ on binary states. Consequently

$$f(y) - \min f = \eta + B + 2\epsilon - \delta_B - \delta_e. \tag{20}$$

Sending both positive deltas to zero approaches the full additive bound. The construction fits the stated model: orthogonal embeddings for u, v and zero embeddings for e, s eliminate FM pair terms; the neural input weights supply the single preactivation. It is a worst-case retained-witness construction, not a lower bound against every optimizer. Additional search or a richer archive could find a better candidate. An exhaustive 16-state script accompanies the algebra.

A.4 Context coverage and improvement events

If d is divisible by b, a uniform random partition places a fixed coordinate pair together with probability $(b - 1)/(d - 1)$. Conditional on the first coordinate, the other $b - 1$ block members are a uniform subset. Across T independent partitions, the miss probability is $(1 - (b - 1)/(d - 1))^T$. For a fixed interaction set E, the union bound is $|E|$ times this expression, capped at one. Unequal block sizes $b_1, \ldots, b_L$ instead give one-pass pair probability

$$\frac{\sum_\ell b_\ell(b_\ell - 1)}{d(d - 1)}. \tag{21}$$

Co-activity is a schedule property, not proof that an interaction is learned or optimized. In particular, $b = 1$ never exposes a pair simultaneously.

For a fixed target x^* and active set A, an independent product context contains x^* in its subcube with probability

$$\pi_A = \prod_{j \notin A} p_j^{x_j^*}(1 - p_j)^{1 - x_j^*}. \tag{22}$$

At $p_j = 1/2$, this is $2^{-(d-|A|)}$. Clipping probabilities only supplies the generally tiny lower bound $\delta^{d-|A|}$. An observed-anchor mixture changes this event for an already available anchor: choosing a specified center and preserving its passive coordinates has probability $(\beta/M)(1 - \rho_c)^{d-b}$. This does not solve the problem of discovering an unknown optimum.

To justify (11), $\mathbb{E}W_A = (b/d)W$ and Markov imply $\mathbb{P}(W_A > \tau) \leq \xi$. Useful blocks have probability at least κ, so the intersection of usefulness and $W_A \leq \tau$ has probability at least $\kappa - \xi$, without assuming independence. Multiply the anchor-preservation probability and the assumed history-wise solver-success bound. On that event, Theorem 1 supplies the displayed improvement. For N trials each succeeding with probability at least p conditional on every previous history, iterated conditional expectation bounds failure by $(1 - p)^N$. A marginal success probability alone is insufficient.

The implemented partition's blocks are dependent. A conservative application counts the first full-sized block of each fresh controller iteration under the stated conditional assumptions; it cannot

simply exponentiate a marginal bound by every block visited. Small residual error also does not imply a large κ or Δ. Those unknown quantities encode whether the restricted neighborhood contains worthwhile true improvements.

A.5 TT CONSTRUCTION, SAMPLING LAW AND RECOVERY

A mixture of $m+1$ product distributions has an explicit nonnegative TT: the first core stores mixture weights times first-coordinate probabilities; middle cores are diagonal matrices of component coordinate probabilities; the final core stores last-coordinate probabilities. Contraction gives (8); summing over every coordinate gives one. The uniform component supplies full support. Rank is at most $m+1$, rather than an assertion that every learned core later has this rank or nonnegativity structure.

For distinct seed set S, each seed receives its own mixture-component contribution, so

$$P_0(S) \geq (1-\omega)(1-\rho)^d. \tag{23}$$

For a designated subset of g seeds, multiply this bound by g/m. If $\rho = c/d \leq 1/2$, then $\log(1-\rho) \geq -2\rho$, giving $(1-\rho)^d \geq e^{-2c}$ and limit e^{-c}. These are exact-seed recovery bounds for a fixed initialization, not claims that all its samples are useful.

More generally, let P be the normalized sampling law and S consist of m distinct points with finite mean negative log-likelihood $\mathcal{L}_S = -m^{-1}\sum_{x\in S}\log P(x)$. Arithmetic-geometric mean gives

$$P(S) \geq me^{-\mathcal{L}_S}, \qquad \mathbb{P}(\text{hit } S \text{ in } K \text{ iid draws}) \geq 1-(1-me^{-\mathcal{L}_S})^K. \tag{24}$$

Zero-probability seeds make the likelihood bound vacuous. Duplicate rows cannot be counted as distinct mass. If a good retained seed is already in $\mathcal{D}$, its true value is already known, but a guarantee for a new query requires an eligible unseen seed and its own score bound.

For the ideal mixture on 0^d and 1^d, take first cores $[1/2, 0]$ or $[0, 1/2]$, middle cores $\mathrm{diag}(1,0)$ or $\mathrm{diag}(0,1)$, and last cores $[1,0]^\top$ or $[0,1]^\top$. Contraction gives half mass on each mode, with every nontrivial unfolding of rank two. The independent marginals are uniform and assign their union only 2^{1-d}. This familiar representation example does not demonstrate that training will learn these modes.

After gradient updates, the reference PROTES implementation can have signed cores. Its sampling law uses sequential normalized absolute conditional contractions. We use the corresponding likelihood and do not interpret it as Born sampling or as simply normalizing signed tensor entries. The nonnegative initialization agrees with that law initially; fixed-distribution iid recovery assertions do not automatically extend to adaptive updates. The recorded float32 failure and separate diagnostic replay document why finite entries alone do not guarantee numerical conditional likelihoods.

A.6 COMPUTABLE QUADRATIC LOWER BOUND

Write $\psi(x) = c + a^\top x + \|V^\top x\|^2/2$, where $a_i = w_i - \|v_i\|^2/2$. Completing a square and minimizing each binary coordinate yields

$$L(\lambda) = c - \tfrac{1}{2}\|\lambda\|^2 + \sum_i \min(0, a_i + (V\lambda)_i) \leq \min_{x\in\{0,1\}^b} \psi(x). \tag{25}$$

For any finite λ, nonnegativity of $\|V^\top x - \lambda\|^2/2$ gives

$$c + a^\top x + \tfrac{1}{2}\|V^\top x\|^2 \geq c - \tfrac{1}{2}\|\lambda\|^2 + (a + V\lambda)^\top x. \tag{26}$$

Minimize each binary coordinate on the right to obtain (25). The maximum over any finite set of evaluated λ remains a lower bound. A relaxed box candidate u can suggest $\lambda = V^\top u$, but its numerical primal objective is not itself a certified lower bound. The implementation always evaluates the explicit dual expression.

Combining $\min_{\mathcal{C}} g \geq L_* + r_-$ with a returned y yields $g(y) - \min_{\mathcal{C}} g \leq g(y) - L_* - r_- \leq \psi(y) - L_* + B$. Under the separate uniform approximation assumption, add 2ϵ for the true conditional gap. Sampling several subcubes does not make their lower bounds a global certificate without established coverage. The 72-cell assay uses saved development tensors and zero new physical calls; all numerical assertions are tolerance-based float64 checks without outward rounding.

B IMPLEMENTATION AND REPRODUCIBILITY

B.1 IMPLEMENTATION, EVALUATION AND DATA DETAILS

Table 4: Frozen default settings. Task-specific overrides are recorded in immutable manifests.

Component	Setting
FM embedding / neural hidden width	8 / 32
Training epochs / early-stop patience	100 / 15
AdamW learning rate / weight decay	10^{-3} / 0.1
Dropout / validation fraction	0.1 / 0.2
Auxiliary quadratic supervision	0 (development evidence mixed)
Controller passes / contexts per block	8 / 1
Annealer reads / sweeps	8 / 8
Local retained candidates / archive capacity	8 / 512
Elites / offspring / genetic generations	24 / 32 / 1
Controller smoothing / clipping floor	0.5 / 0.05
Observed anchors / anchor-mixture weight	8 / 0.5
Expected flips around an anchor	1
TT rank / maximum seed centers	8 / 7
Uniform TT mixture / expected seed flips	0.1 / 1
TT surrogate scores / batch / elites	512 / 64 / 8
TT gradient steps / learning rate	1 / 0.01

Outer-loop refitting and label retention. At round t, fit a fresh DeepFM model g_t from the entire accumulated dataset $\mathcal{D}_t$, with a newly seeded 80/20 training/validation split and validation-based early stopping. Model weights are not warm-started. Freeze this model while the seeder and PROTES search it. If $\mathcal{U}_t$ denotes the distinct, previously unevaluated candidates scored by PROTES, select the lowest-g_t candidates in $\mathcal{U}_t$ up to the remaining batch/budget limit. The final surrogate-method protocols use 32 true queries per round. Query f only for this selected batch B_t and set

$$\mathcal{D}_{t+1} = \mathcal{D}_t \cup \{(x, f(x)) : x \in B_t\}. \tag{27}$$

All returned labels are retained, including poor designs; there is no top-only training filter and no use of surrogate predictions as training labels. If too few eligible candidates exist, the recorded uniform fallback fills the batch. With 128 initial labels and a 256-call budget, the usual high-dimensional run has four search/refit rounds. The terminal batch is recorded even though no further model is needed after the budget is exhausted. Initialization, inner surrogate evaluations, and true evaluations therefore have distinct roles and counters.

The matched full-QUBO seeder uses the factorized FM solver. Random partitions change at each controller pass. The archive includes complete candidates before selection; its capacity limits storage, and the best full-score candidate is retained. Correction construction, annealing, fitting and TT phases are timed separately. Tensor storage remains $O(dr^2)$ and a sequential TT sample requires $O(dr^2)$ work. The flattened hidden layer has $O(dkH)$ weights, so a fixed small label budget does not guarantee learnability.

BODi uses the authors' binary dictionary kernel and acquisition optimizer at commit `aa507d34a96407b647bf808375b5e162ddf10664`. The setup uses 64 dictionary vectors, a Matérn-5/2 ARD kernel, standardized maximization targets equal to negative loss, expected improvement, 2,000 random acquisition candidates, 20 local restarts, at most 50 neighbor-search steps, and incumbent spray points. It refits after each sequential observation. Binary Hamming distances are evaluated by an equivalent matrix-product identity; joint posterior marginals evaluate the same one-point EI without replicating training data. Independent tests check encoding, EI values and proposal agreement. The adapter updates the GP fitting API without changing to log-EI or adding outcome standardization twice. Upstream source is fetched and hash-verified rather than redistributed without a license.

Some original cloud sessions reached their time cap or were cancelled. Recovery executes only never-started frozen tasks, preserving completed and interrupted attempts. Certain recovery sessions use a different CPU/GPU host; manifests record this migration. Shared-start, source, budget and numerical audits still apply, but recovered timings are not used as causal speed comparisons across hosts, and numerical trajectories need not be bitwise identical. Provider evidence, rather than a stale local progress flag, identifies interrupted runs.

CEM, GA and direct PROTES optimize the true objective with the same charged initial designs and budget. Direct PROTES receives the shared starts through the reference implementation's external-sampling interface and updates its TT on their true values; initialization is not discarded before learning. Its batch size is 32 for these final protocols, rank eight, learning rate 0.01, eight elites and one gradient step per batch. BODi's original initialization is replaced by these shared starts. Repeated proposals are not new objective information: duplicates are handled by the declared outer-loop rule and any uniform fallback is recorded. Query efficiency and wall-clock efficiency are distinct; sequential BODi and batched surrogate methods are explicitly identified.

The final controlled studies use five instance seeds at each intermediate dimension and both higher-order settings, with each instance shared across methods. The separate scaling study uses five other instances and nine methods. Film uses ten seeds; each optical study uses five. Paired bootstrap intervals are descriptive, with 10,000 resamples from a fixed analysis seed. Reusing instance identities across conditions does not create independent observations. An exact two-sided sign test cannot reach 0.05 with only five non-tied pairs; no such significance is asserted.

All comparisons use the frozen settings. Their results do not establish competitiveness after separate extensive hyperparameter optimization for every baseline. Matched tuning effort and held-out model-selection tasks remain a future evaluation requirement.

The optical structured-initialization comparison covers anchored FM, A2+, A3+, GA and BODi. Component controls include anchored raw A3, corrected A3 without anchor context access, a fixed controller, 16-bit and full-sized blocks, no genetics, no TT updates and a smaller archive. Their frozen settings and any unfinished outcomes must remain in the final inventory. Incomplete cohorts are excluded from performance tables but remain explicitly identified in the status account and artifact.

Prospective fresh-seed follow-up. After observing the original 1,024-variable optical result, we froze a separate follow-up with seeds 1401–1410 and unchanged algorithms/settings. Six methods—FMQA, A1, A2+, A3+, GA and BODi—are included in two fixed five-seed jobs. The primary comparison is A3+ minus A2+ after 10 nm reevaluation; all completed secondary comparisons and incomplete statuses must be reported. This is a prospectively fixed replication of a selected finding, not an originally blinded task selection. A later user-requested extension adds direct PROTES on all ten fresh seeds, frozen before inspecting any follow-up outcomes; the original primary comparison is unchanged. Cross-job initial arrays are checked before paired reporting. These follow-up results are not pooled with the original five seeds or used to retune the method. The direct-PROTES extension uses a GPU-host session because CPU slots were occupied; PROTES and the optical solver still execute on CPU. The recorded host amendment changes no algorithm, seeds or budgets, but prevents a controlled cross-host timing interpretation.

Prospective label-scarce scaling follow-up. A separate 36-task protocol, frozen after the original outcomes, tests 20,000 binary variables with 100 true-objective evaluations: 40 shared initial observations followed by three batches of 20. Three fresh paired seeds (1501–1503) compare FMQA, anchored FMQA, A1, A2+, A3+, random search, CEM, GA, direct PROTES and sequential BODi. Inner PROTES search uses 4,096 full-surrogate evaluations per round; additional A1 and A3+ controls use 512. The primary mechanistic contrast is high- versus low-budget A3+, with all other comparisons secondary. True calls, surrogate calls and full computation are counted separately. The unchanged optimizer is evaluated on constructed polynomial objectives, whose true evaluations are inexpensive; this is not a large photonic experiment or evidence of an expensive-oracle speedup. Three-seed cohorts are descriptive, unfinished endpoints and failures remain visible, and outcomes are not pooled with the 256-call study. Concurrent local execution shares resources with existing recovery jobs, limiting causal timing comparisons.

B.1.1 PHYSICAL DEFINITIONS AND PROVENANCE

For each wavelength let $P_{\lambda j} = |S_{\lambda,\mathrm{in},j}|^2$. The WDM loss is

$$\tfrac{1}{2}\left[(1 - P_{1270,1})^2 + P_{1270,2}^2 + (1 - P_{1290,2})^2 + P_{1290,1}^2\right]. \tag{28}$$

The design controls expand to the simulator's physical region; increasing control count changes actual binary material decisions. The model uses nondispersive dielectric values and two-dimensional electromagnetic fields. Fine-grid validation reconstructs the same selected control map, keeps physical port offsets fixed, checks sparse-system relative residuals, and recomputes loss from saved complex scattering. It does not assume that a converged linear solve establishes continuum or fabrication accuracy. Earlier coarse-grid improvements that failed refinement are excluded from final success claims, but retained as development observations.

The planar pilot alternates allowed SiO_2/TiO_2 material choices at fixed thickness, with air ambient and a semi-infinite silica substrate. TiO_2 data are linearly interpolated in wavelength, including extinction coefficient, without extrapolation. The Malitson formula is used within its original wavelength interval. The data files retain CC0 database headers and scientific attribution. TiO_2 source data describe a 350 nm ALD film, whereas modeled layers are 75 nm; this transfer is a limitation. The material arrays and SHA-256 checksums are part of the artifact.

Additional physical illustrations use a prespecified lower-median A3+ run, with stable seed tie-breaking; its best charged design is compared to its common initial best. WDM selection ranks the separately evaluated 10 nm terminal losses while retaining each design selected on the 20 nm grid. Film selection ranks terminal optimization loss. Extra spectra do not select another design or feed back to optimization. WDM spectra use 31 frequencies per design, while film spectra use 301; this work is accounted separately from ordinary objective calls. The included figures use completed groups that passed the evidence audit.

B.2 DETAILED CONDITIONAL SEED ENGINE

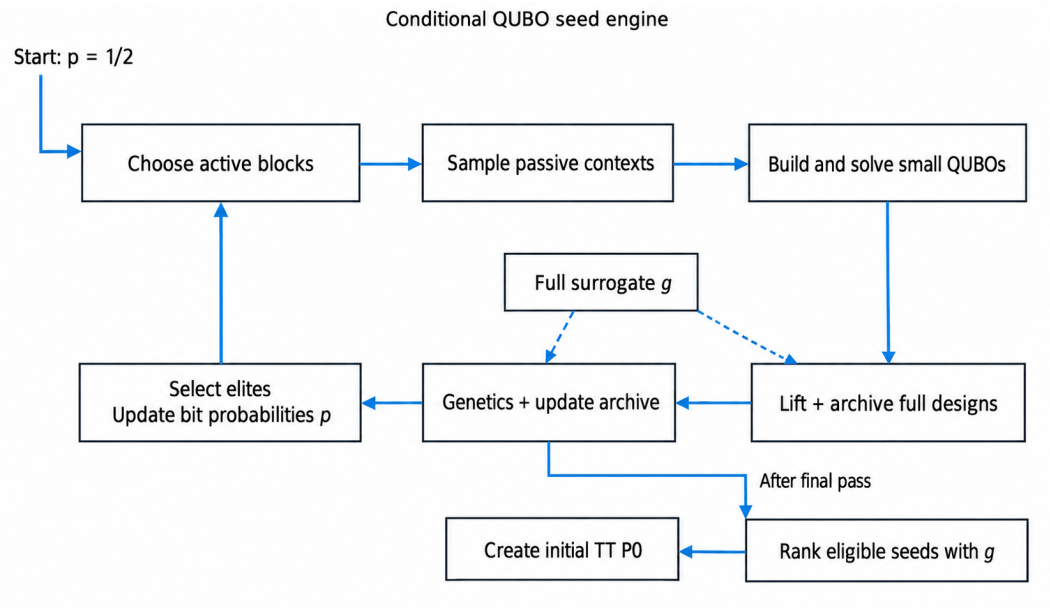

Figure 4: Conditional seed generation. Each pass forms active blocks from a fresh permutation and samples passive contexts, with an observed-anchor mixture in A3+. Every block/context solve is lifted and archived before the pass's genetic stage; offspring also update that archive. Full-g-ranked elites update the product bit probabilities p, which drive the next pass. After the final pass, eligible complete seeds from the updated archive are ranked to create P_0. The product probabilities and TT distribution are distinct. Dashed arrows denote full-surrogate scoring; no inner step queries the true objective.

B.3 ARCHITECTURE CONTROLS

Table 5: Architecture controls with the same DeepFM predictor and full-surrogate PROTES search. TTC ranks complete seeds using g and constructs the initial TT. A2+/A3+ add the same affine-residual correction and observed-anchor access. A3 variants alone use context updates, complete-vector archiving and genetics. The TT always spans all d variables.

Variant	QUBO size	Correction	Anchors	TT initialization
A1	None	None	No	Default
A2	d	None	No	TTC
A3	At most b	None	No	TTC
A2+	d	Affine residual	Yes	TTC
Trinity (A3+)	At most b	Affine residual	Yes	TTC

C EXTENDED EXPERIMENTAL RESULTS

C.1 COMPLETED SCALING CURVES AND REPRODUCIBILITY CHECKS

Table 6: Completed final scaling study: terminal loss, mean $\pm$ standard deviation, 256 charged queries including 128 initial designs. Lower is better; each displayed cell has five completed runs. Raw A3 is omitted from this paired two-size table because its 10,000-variable cohort is incomplete; its complete 5,000-variable result appears in the next figure.

Method	5,000 variables	10,000 variables
FMQA + anchors	$\mathbf{0.6819 \pm 0.0014}$	0.6891 ± 0.0007
A1: DeepFM + PROTES	0.6885 ± 0.0005	0.6894 ± 0.0011
A2: full seeds	0.6866 ± 0.0025	0.6898 ± 0.0014
A2+: corrected full	0.6860 ± 0.0018	0.6898 ± 0.0007
A3+: corrected conditional	0.6853 ± 0.0019	0.6892 ± 0.0017
CEM	0.6826 ± 0.0012	$\mathbf{0.6871 \pm 0.0018}$
Genetic search	0.6834 ± 0.0027	0.6880 ± 0.0007
Direct PROTES	0.6874 ± 0.0012	0.6901 ± 0.0010

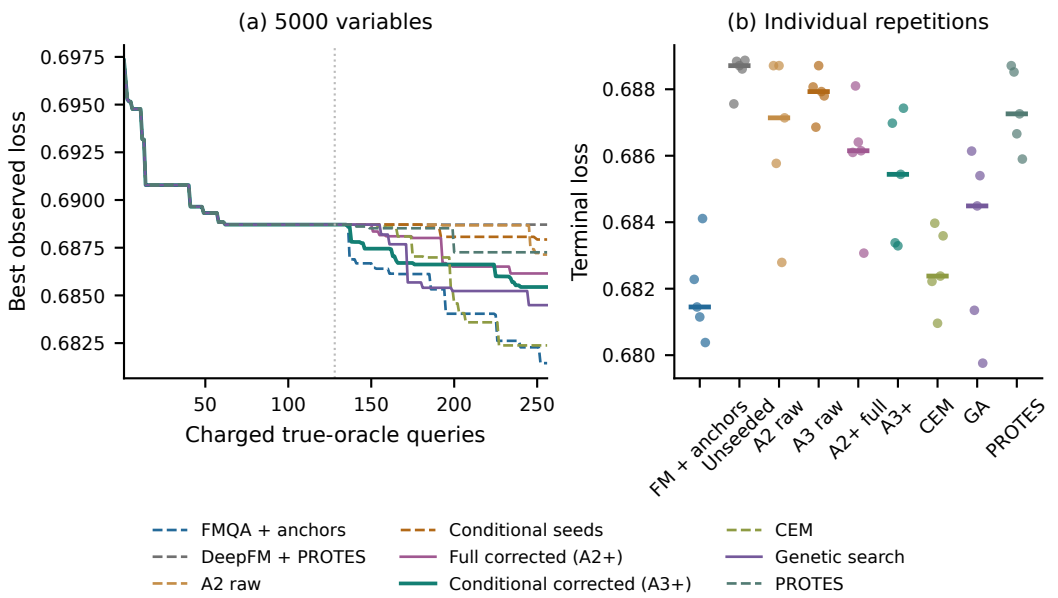

Figure 5: All nine methods on the completed 5,000-variable study. Curves summarize best charged values; points show individual final outcomes. These are five paired instances under a fixed 256-call budget, not evidence of dominance across arbitrary large problems.

The raw A3 cohort at 10,000 variables has four of five completed runs; the failed run stopped after initialization in a float32 TT likelihood diagnostic. Its partial aggregate and failed prefix are excluded from the performance display. A float64 diagnostic replay did not replace the frozen outcome. Full logs and all completed observations remain in the artifact.

The source audit checks immutable manifests, source hashes, task identity, unique charged queries and common initial data across workers. Optical validation checks design identity, scattering-derived loss and numerical residuals. Mutation tests deliberately corrupt shared starts and physical data to ensure the audit rejects them. An independently extracted staging artifact passed 44 tests with the optional BODi class initially skipped; after its documented fetcher verified the pinned upstream source, all three BODi checks passed separately in that extracted artifact. These checks used the recorded host environment, not a clean dependency installation, and are software verification rather than scientific replications. The accompanying artifact records the captured completion inventory and retains all unavailable or interrupted endpoints.

C.2 Complete final film outcomes

Table 7: All 13 methods, ten paired final seeds, 256 total objective calls including 64 common initial designs. Lower loss is better.

Method	n/N	Loss (mean ± SD)	Time (s), median
FM	10/10	0.0913 ± 0.0268	6.9
FMQA + anchors	10/10	0.0902 ± 0.0315	6.8
FM/TT	10/10	0.0807 ± 0.0178	32.5
DeepFM + PROTES	10/10	0.0844 ± 0.0180	32.2
A2 raw	10/10	0.0856 ± 0.0312	35.1
Conditional seeds	10/10	0.0821 ± 0.0262	33.7
Full corrected (A2+)	10/10	0.0719 ± 0.0246	33.8
Conditional corrected (A3+)	10/10	0.0715 ± 0.0263	32.5
Random	10/10	0.0982 ± 0.0277	5.9
CEM	10/10	0.0707 ± 0.0227	5.9
Genetic search	10/10	0.0560 ± 0.0076	5.7
PROTES	10/10	0.0826 ± 0.0273	13.7
BODi	10/10	$\mathbf{0.0136 \pm 0.0063}$	639.6

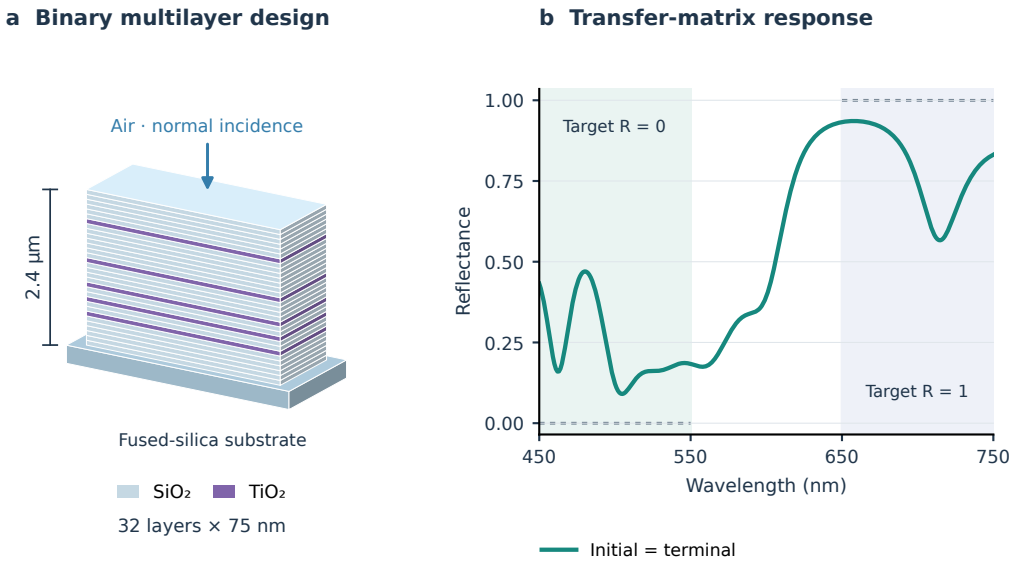

Figure 6: Prespecified lower-median A3+ example, selected by terminal objective with a stable seed tie-break. Here its terminal best is the common initial best: the identical designs and overlapping spectra are retained, rather than choosing a more favorable run. Layers are ordered from air to the semi-infinite fused-silica substrate; lateral dimensions show an illustrative finite window into the planar model. Colored spectral regions mark objective bands; the unshaded middle interval is unconstrained. Two additional 301-frequency spectra were evaluated without feedback.

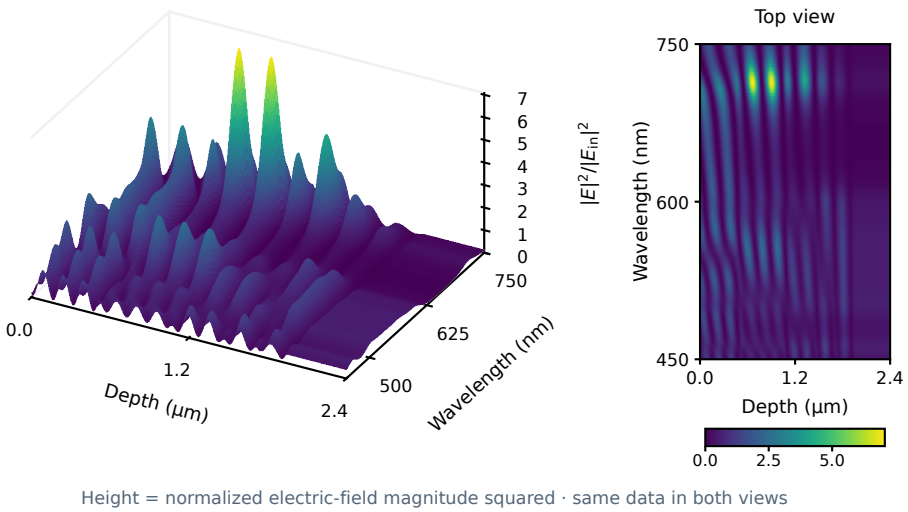

Figure 7: Actual depth-resolved response of the same prespecified median film design. Height denotes $|E|^2/|E_{\text{in}}|^2$ over physical depth and wavelength, with the same data shown from above; it is not a third spatial simulation coordinate. The initial and terminal design are identical. The planar transfer-matrix calculation follows Byrnes (2016); 151 additional frequency solves at 321 depth positions were recorded without optimization feedback. These are additional to the 602 spectral solves above. Reflectance agrees with the stored spectrum, maximum tangential-field interface discontinuity is 8.96×10^{-16}, and maximum boundary-power mismatch is 4.28×10^{-15}. Field enhancement does not establish optimization superiority.

C.3 Completed controlled 512-variable study

All 130 prespecified runs completed: thirteen methods, two interaction settings, and five shared instances per setting. Table 8 includes every method. BODi beats A3+ in each of the ten pairs. The mean A3+ minus BODi differences are 0.04027 at $\nu = 0$ and 0.03254 at $\nu = 0.5$, with descriptive paired 95% bootstrap intervals $[0.02703, 0.05219]$ and $[0.02391, 0.04273]$, respectively. These intervals are unadjusted for multiple comparisons.

Table 8: Completed controlled 512-bit study: terminal loss, mean ± sample standard deviation, five paired instances and 256 true calls including 128 shared initial observations. Lower is better.

Method	$\nu = 0$	$\nu = 0.5$
FMQA-style	0.45391 ± 0.01015	0.60553 ± 0.01196
FMQA + anchors	0.45258 ± 0.00557	0.60732 ± 0.01273
FM + PROTES	0.46258 ± 0.00316	0.66295 ± 0.00579
A1: DeepFM + PROTES	0.46437 ± 0.00544	0.66164 ± 0.00647
A2: full seeds	0.46320 ± 0.00630	0.60840 ± 0.01663
A3: conditional seeds	0.46543 ± 0.00562	0.64533 ± 0.00859
A2+: corrected full	0.45918 ± 0.00360	0.60881 ± 0.01509
A3+: corrected conditional	0.46301 ± 0.00642	0.62510 ± 0.01034
Random search	0.46555 ± 0.00579	0.66268 ± 0.00655
CEM	0.46219 ± 0.00765	0.63986 ± 0.00875
Genetic search	0.45578 ± 0.00285	0.64941 ± 0.00244
Direct PROTES	0.46461 ± 0.00487	0.66314 ± 0.00573
BODi	$\mathbf{0.42273 \pm 0.01176}$	$\mathbf{0.59256 \pm 0.00471}$

At $\nu = 0.5$, A3+ beats A1 in all five pairs with mean difference -0.03654 (interval $[-0.04627, -0.02682]$), and GA with difference -0.02432 (interval $[-0.03111, -0.01545]$). Its comparison against corrected full-seeding A2+ is mixed: two wins and three losses, mean difference $+0.01629$ (interval $[-0.00137, 0.03395]$). At $\nu = 0$, A3+ has no paired win against A2+ (two ties, three losses). The observations distinguish gains from surrogate-guided seeding from a general benefit of bounded decomposition. Original and recovered executions share the frozen protocol but may have different host characteristics; no causal timing comparison is inferred.

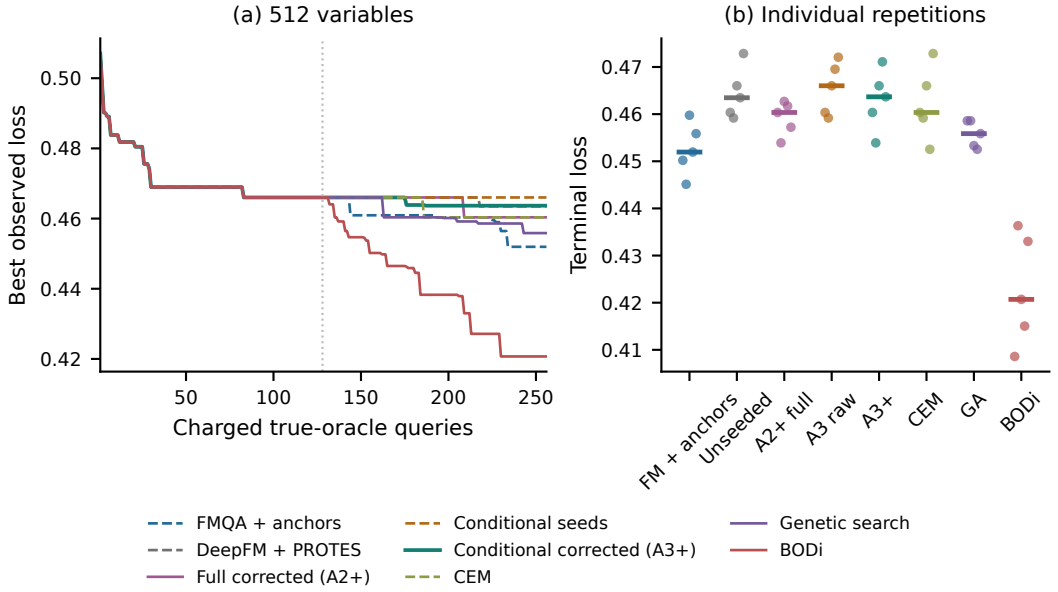

Figure 8: Controlled 512-bit study, $\nu = 0$: median incumbent traces and all five individual terminal observations for the fixed plotting subset. The dotted line marks the end of shared initialization. The full thirteen-method inventory appears in Table 8.

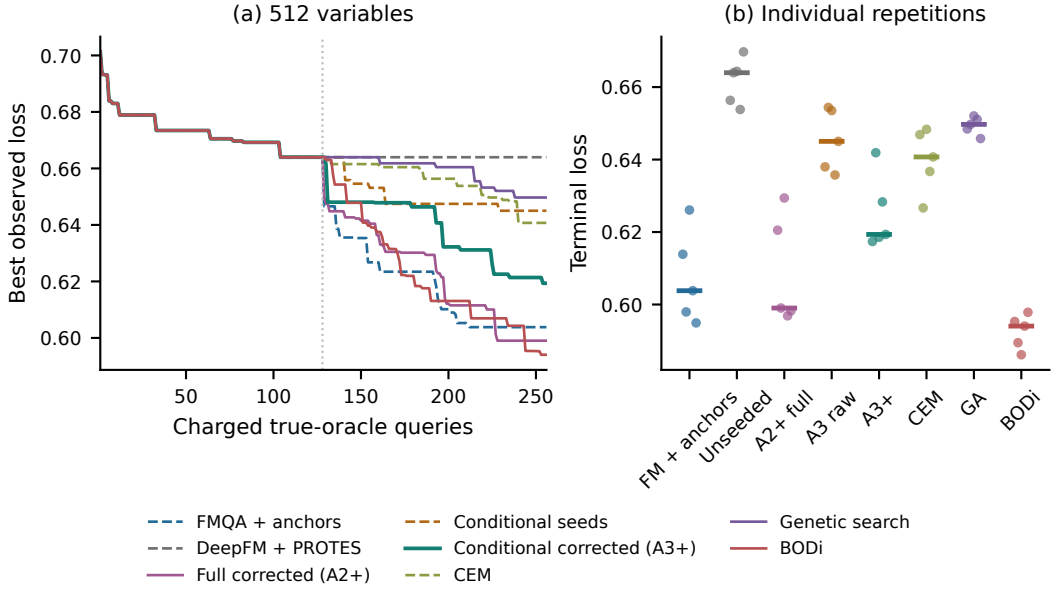

Figure 9: Controlled 512-bit study, $\nu = 0.5$, using the same method subset and display conventions. Better outcomes against unseeded A1 do not imply superiority over corrected full seeding or BODi.

C.4 COMPLETED CONTROLLED 1,024-VARIABLE STUDY

All 130 planned endpoints are complete and audited, including the unchanged recovered tasks. The two conditions show different decomposition effects: for $\nu = 0$, A3+ beats corrected full seeding A2+ on all five instances, with mean paired difference -0.01014 and descriptive 95% bootstrap interval $[-0.01604, -0.00654]$. At $\nu = 0.5$, it loses all five pairs to A2+, with mean difference $+0.02697$ and interval $[0.01717, 0.03507]$. A3+ beats A1 and direct PROTES on every pair in both conditions, but BODi beats A3+ on all ten pairs. The full inventory below preserves both outcomes; intervals are unadjusted and descriptive.

Table 9: All thirteen methods at 1,024 variables: mean $\pm$ sample standard deviation, five paired instances per condition, 256 true calls including 128 shared starts. Lower is better.

Method	$\nu = 0$	$\nu = 0.5$
FMQA-style	0.47604 ± 0.00363	0.61435 ± 0.01353
FMQA + anchors	0.47041 ± 0.00761	$\mathbf{0.60939 \pm 0.00569}$
FM + PROTES	0.47678 ± 0.00343	0.67316 ± 0.00194
A1: DeepFM + PROTES	0.47727 ± 0.00361	0.67188 ± 0.00393
A2: full seeds	0.47541 ± 0.00746	0.63242 ± 0.00808
A3: conditional seeds	0.47762 ± 0.00414	0.66541 ± 0.00310
A2+: corrected full	0.47734 ± 0.00376	0.63050 ± 0.01026
A3+: corrected conditional	0.46721 ± 0.00748	0.65747 ± 0.00496
Random search	0.47529 ± 0.00366	0.67456 ± 0.00182
CEM	0.47668 ± 0.00355	0.65417 ± 0.01115
Genetic search	0.47150 ± 0.00378	0.66275 ± 0.00627
Direct PROTES	0.47598 ± 0.00286	0.67385 ± 0.00405
BODi	$\mathbf{0.45043 \pm 0.00610}$	0.61968 ± 0.00340

C.5 COMPLETED CONTROLLED 2,048-VARIABLE COMPARISONS

The frozen controlled study has completed all five repetitions for each of twelve methods in both polynomial conditions (120 runs). Table 10 includes those twelve complete cohorts. The thirteenth method, BODi, has ten incomplete planned endpoints: two interrupted attempts and eight unavailable tasks. No partial BODi prefix is used as a terminal result. This is a completed architecture comparison within an incomplete all-method study.

Table 10: Controlled 2,048-bit terminal loss, mean $\pm$ sample standard deviation over five paired instances, 256 true calls including 128 shared initial observations. Lower is better. Only complete cohorts are displayed; the excluded BODi cohort receives no aggregate mean or rank.

Method	$\nu = 0$	$\nu = 0.5$
FMQA-style	0.48174 ± 0.00373	$\mathbf{0.64772 \pm 0.00611}$
FMQA + anchors	0.47966 ± 0.00641	0.65048 ± 0.00568
FM + PROTES	0.48321 ± 0.00142	0.67982 ± 0.00129
A1: DeepFM + PROTES	0.48492 ± 0.00189	0.68064 ± 0.00268
A2: full seeds	0.48469 ± 0.00207	0.66359 ± 0.00603
A3: conditional seeds	0.48470 ± 0.00184	0.67874 ± 0.00223
A2+: corrected full	0.48281 ± 0.00194	0.66357 ± 0.00569
A3+: corrected conditional	$\mathbf{0.47950 \pm 0.00337}$	0.67330 ± 0.00293
Random search	0.48453 ± 0.00233	0.68144 ± 0.00272
CEM	0.48321 ± 0.00061	0.67125 ± 0.00451
Genetic search	0.48065 ± 0.00234	0.67550 ± 0.00179
Direct PROTES	0.48294 ± 0.00203	0.67971 ± 0.00328

A3+ improves over unseeded A1 on all five instances in both conditions. For $\nu = 0$, its mean paired difference is -0.00542 with descriptive 95% bootstrap interval $[-0.00699, -0.00349]$. However,

its difference against corrected full seeding A2+ is -0.00331 with interval $[-0.00579, 0.00057]$; the interval against GA also spans zero. At $\nu = 0.5$, A3+ loses to A2+ on all five instances, with mean difference $+0.00972$ and interval $[0.00534, 0.01426]$. FMQA-style search has the lowest completed mean in this condition. Thus a benefit over unseeded nonlinear search does not establish a decomposition benefit or literature superiority. The intervals are descriptive and unadjusted for multiple comparisons. Some recovered runs use a different host, as disclosed in Appendix B.1; these observations do not establish wall-clock advantage.

C.6 Completed 1,024-variable photonic study

Every planned method completed five shared-seed runs, with 256 charged calls including 128 common initial designs. Each terminal design selected on the 20 nm grid was independently reevaluated at 10 nm without optimization feedback. Mean and sample standard deviation are reported below; lower is better.

Table 11: All thirteen methods on the 1,024-variable WDM task. The complete five-seed cohort is retained for every method.

Method	20 nm selection grid	10 nm reevaluation
FMQA	0.7602 ± 0.0103	0.7729 ± 0.0165
FMQA + anchors	0.7549 ± 0.0165	0.7640 ± 0.0087
FM + PROTES	0.7677 ± 0.0140	0.7857 ± 0.0238
A1: DeepFM + PROTES	0.7677 ± 0.0140	0.7857 ± 0.0238
A2: full seeds	0.7677 ± 0.0140	0.7857 ± 0.0238
A3: conditional seeds	0.7677 ± 0.0140	0.7857 ± 0.0238
A2+: corrected full	0.7677 ± 0.0140	0.7857 ± 0.0238
A3+: corrected conditional	0.7234 ± 0.0248	0.7391 ± 0.0237
Random search	0.7677 ± 0.0140	0.7857 ± 0.0238
CEM	0.7506 ± 0.0153	0.7582 ± 0.0316
Genetic search	0.7180 ± 0.0220	0.7342 ± 0.0345
Direct PROTES	0.7677 ± 0.0140	0.7857 ± 0.0238
BODi	$\mathbf{0.6307 \pm 0.0925}$	$\mathbf{0.6594 \pm 0.0992}$

A3+ versus A2+ has mean paired refined-loss difference -0.04666, descriptive 95% bootstrap interval $[-0.07552, -0.01775]$, and four wins/one tie. Against FMQA the difference is -0.03383, interval $[-0.05486, -0.01376]$, also four wins/one tie. The mean-loss reductions are 5.9% and 4.4%; they are ratios of cohort means, not averages of per-run percentage gains. Anchored FMQA has mean 0.76401 and is also worse in four pairs with one tie. BODi remains stronger in four pairs and has mean 0.65941; GA has mean 0.73423 and a paired interval against A3+ spanning zero. Five seeds and unadjusted descriptive intervals do not establish broad statistical superiority. Recovery host changes preclude interpreting cross-method elapsed times as a controlled hardware speed comparison.

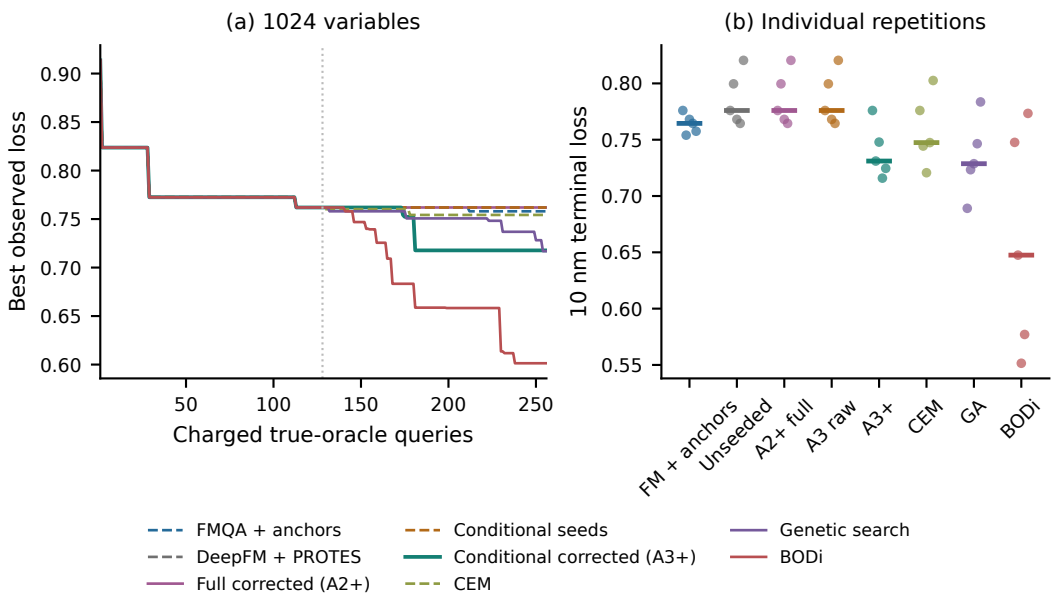

Figure 10: Optimization-grid incumbent traces and all five individual endpoints for the fixed plotting subset. Full method coverage and fine-grid values appear in the table above. A3+ improves over the full-seeding architecture; the literature baseline comparison remains mixed.

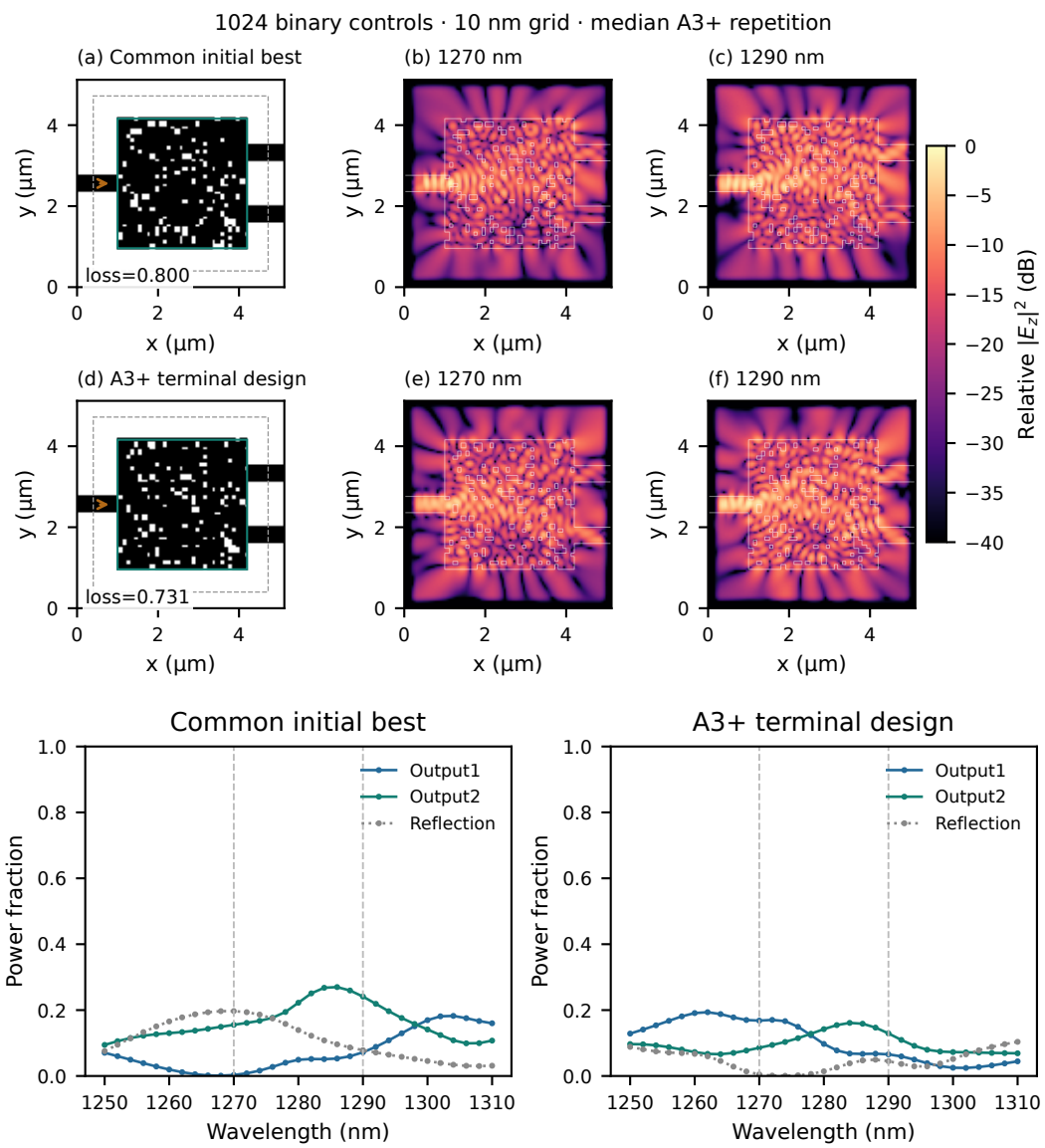

Figure 11: Physical response of the prespecified lower-median A3+ repetition (seed 1002), comparing the common initial best with the terminal design shown in Figure 1. Top: actual permittivity layouts and simulated field intensity at the two objective wavelengths. Each wavelength shares its intensity normalization across designs. Bottom: actual 2 nm-spaced spectral samples, joined linearly; dashed lines mark the objective wavelengths. These post-selection visualizations required 62 additional wavelength solves, separately recorded and never fed back to optimization. Recomputed target losses agree with saved fine-grid evaluations to within 6.4×10^{-9}. The model is two-dimensional and nondispersive; this example is neither a fabricated-device measurement nor evidence of broadband optimality.

C.7 Other photonic dimensions and complete component controls

The 512-variable study completed all 65 runs and independent refinements. A3+ has mean refined loss 0.70872 versus 0.74625 for A2+, with four wins and one tie. Its comparison with FMQA is mixed: two wins and three losses, mean difference -0.00388 and descriptive 95% interval $[-0.04671, 0.03868]$. BODi wins all five pairs. At 2,048 variables, 64/65 outcomes and refinements are complete; one BODi run was interrupted by the original provider session. The remaining recovered tasks retain their frozen settings. A3+ improves over A2+ in four pairs and loses one,

while CEM and GA have lower means. We withhold the incomplete BODi cohort mean. All tables highlight the lowest unrounded mean among complete displayed cohorts, with ties at tolerance 10^{-12}; boldface is not a significance test.

Table 12: All thirteen methods at 512 optical variables, five paired seeds: mean ± sample SD at selection and independent reevaluation. Lower is better.

Method	20 nm selection grid	10 nm reevaluation
FMQA	0.7009 ± 0.0427	0.7126 ± 0.0387
FMQA + anchors	0.7096 ± 0.0400	0.7301 ± 0.0252
FM + PROTES	0.7285 ± 0.0262	0.7489 ± 0.0227
A1: DeepFM + PROTES	0.7334 ± 0.0238	0.7500 ± 0.0235
A2: full seeds	0.7270 ± 0.0254	0.7335 ± 0.0268
A3: conditional seeds	0.7454 ± 0.0358	0.7545 ± 0.0288
A2+: corrected full	0.7211 ± 0.0265	0.7463 ± 0.0338
A3+: corrected conditional	0.7103 ± 0.0239	0.7087 ± 0.0406
Random search	0.7431 ± 0.0322	0.7543 ± 0.0285
CEM	0.6945 ± 0.0365	0.7044 ± 0.0477
Genetic search	0.6817 ± 0.0211	0.6994 ± 0.0370
Direct PROTES	0.7325 ± 0.0238	0.7420 ± 0.0234
BODi	$\mathbf{0.5299 \pm 0.0442}$	$\mathbf{0.5527 \pm 0.0593}$

Table 13: The twelve complete five-seed cohorts at 2,048 optical variables. BODi is excluded because one of its five endpoints is unavailable; the four completed outcomes remain archived.

Method	20 nm selection grid	10 nm reevaluation
FMQA	0.7867 ± 0.0305	0.7993 ± 0.0189
FMQA + anchors	0.7439 ± 0.0249	0.7458 ± 0.0342
FM + PROTES	0.7874 ± 0.0310	0.7991 ± 0.0240
A1: DeepFM + PROTES	0.7874 ± 0.0310	0.7991 ± 0.0240
A2: full seeds	0.7827 ± 0.0254	0.7945 ± 0.0146
A3: conditional seeds	0.7827 ± 0.0255	0.7893 ± 0.0072
A2+: corrected full	0.7874 ± 0.0310	0.7991 ± 0.0240
A3+: corrected conditional	0.7406 ± 0.0354	0.7610 ± 0.0579
Random search	0.7874 ± 0.0310	0.7991 ± 0.0240
CEM	0.7242 ± 0.0369	$\mathbf{0.7330 \pm 0.0474}$
Genetic search	$\mathbf{0.7114 \pm 0.0200}$	0.7335 ± 0.0191
Direct PROTES	0.7687 ± 0.0217	0.7722 ± 0.0175

C.7.1 Ablations and their proper references

The eight frozen ablations contribute 40 completed runs with independent 10 nm reevaluation. Forty-five cross-job checks confirm identical 128-design initial arrays and numerically matching labels against the two five-seed references. Most interventions modify A3+; the frozen fixed-controller implementation instead modifies raw A3, so its component comparison uses raw A3. No claim that controller freezing alone explains its difference from A3+ is warranted.

Table 14: Every frozen ablation and both references. Mean ± sample SD over five shared seeds; 256 true evaluations. Bold denotes the lowest complete mean per grid. Main Table 2 supplies the appropriate reference for each intervention.

Setting	20 nm loss	10 nm loss
A3+ (reference)	0.71027 ± 0.02389	0.70872 ± 0.04061
No neural correction	0.72172 ± 0.05426	0.73580 ± 0.05562
No observed anchors	0.74546 ± 0.03592	0.75961 ± 0.03737
No genetics	0.69810 ± 0.06825	$\mathbf{0.69364 \pm 0.05628}$
No TT updates	0.72997 ± 0.04069	0.74621 ± 0.04492
Archive capacity 32	0.70970 ± 0.02753	0.71474 ± 0.03363
Block size 16	$\mathbf{0.69328 \pm 0.06953}$	0.70250 ± 0.08231
Block size 512	0.70556 ± 0.05676	0.71808 ± 0.06511
Raw A3 (reference)	0.74537 ± 0.03578	0.75450 ± 0.02885
Fixed controller	0.72727 ± 0.02744	0.73894 ± 0.02683

For variant minus A3+, removing anchors gives mean refined-loss difference $+0.05089$, descriptive 95% interval $[0.02017, 0.08160]$, with four losses and one tie. Removing TT updates gives $+0.03749$, interval $[0.00095, 0.07931]$, with four losses and one win. Removing genetics gives -0.01507, interval $[-0.05177, 0.01614]$, with three wins and two losses. Its lower mean does not prove a uniform improvement, but prevents attributing a demonstrated benefit to genetics. Block size 16, a full 512-bit block, archive capacity 32, and removing the neural correction each have intervals spanning zero against A3+. These comparisons are descriptive and unadjusted; changing block size changes both the restriction and the amount of solver work. All signed seed differences are retained in the accompanying JSON, and no post-outcome setting replaces the frozen primary method.

C.8 Physical interpretation of the optical objective

Table 15 is a post hoc decomposition of saved 10 nm scattering arrays, with no new simulations or selection feedback. For each wavelength, $T_{\mathrm{desired}} = |S_{\mathrm{desired}}|^2$, $T_{\mathrm{cross}} = |S_{\mathrm{cross}}|^2$, and insertion loss is $-10 \log_{10} T_{\mathrm{desired}}$. The table averages each metric over the same five completed runs; in particular, mean insertion loss is the mean of the logarithms, not the logarithm of mean transmission. The objective is the wavelength mean of $(1 - T_{\mathrm{desired}})^2 + T_{\mathrm{cross}}^2$. These quantities describe the modeled device, not fabrication yield or measured hardware. Gradient-based inverse design has access to derivatives outside the black-box information setting studied here.

Table 15: Physical outcomes for every completed 1,024-variable method at 256 true calls, independently reevaluated at 10 nm. Transmission is a power fraction (higher desired and lower cross transmission are better); lower insertion loss is better. Bold marks the best mean per column, without a significance claim.

	Desired transmission		Cross transmission		Insertion loss (dB)	
Method	1270 nm	1290 nm	1270 nm	1290 nm	1270 nm	1290 nm
FMQA	0.1007	0.1578	**0.0724**	0.0859	12.6297	8.5494
FM+anc.	0.1139	0.1561	0.0817	0.0912	11.4537	8.6825
A1	0.1151	0.1318	0.0759	0.1173	12.0257	9.5136
A2	0.1151	0.1318	0.0759	0.1173	12.0257	9.5136
A2+	0.1151	0.1318	0.0759	0.1173	12.0257	9.5136
A3	0.1151	0.1318	0.0759	0.1173	12.0257	9.5136
T-D	0.1587	0.1352	0.0997	0.0885	8.0210	8.7610
BODi	**0.2391**	0.1703	0.0977	0.1245	**6.8047**	7.9620
CEM	0.0713	0.2105	0.1268	0.0498	12.4640	7.1332
GA	0.0803	**0.2225**	0.1023	**0.0399**	11.4510	**6.6239**
FM + PROTES	0.1151	0.1318	0.0759	0.1173	12.0257	9.5136
PROTES	0.1151	0.1318	0.0759	0.1173	12.0257	9.5136
Rand.	0.1151	0.1318	0.0759	0.1173	12.0257	9.5136

C.9 Excluded incomplete studies and numerical limitations

No incomplete-cohort means or failed prefixes are used as terminal comparisons. In the 20,000-variable, 100-label follow-up, 30/36 runs completed, but three A3+ runs failed in the TT conditional-likelihood diagnostic; a BODi interruption and two unstarted tasks account for the rest. Neither primary high/low-budget A3+ cohort is complete, so the intended contrast is unavailable. A float64 replay of stored cores resolved nonfinite diagnostic likelihoods without new true calls, but was not a completed repaired-method experiment. Stable 20,000-bit optimization is therefore not established.

In the separately frozen 70-run fresh-seed optical study, only the ten direct-PROTES endpoints and refinements were available at cutoff. No paired A3+ versus A2+ replication can be reported. The 25-run structured-start study also had no available final endpoints. These unfinished comparisons are omitted from results displays; all task identities, completed individual observations, settings, failures and unavailable statuses remain in the artifact. They are not counted as losses, pooled with original cohorts or replaced by favorable development observations.

C.10 Ridge–DeepFM predictor: development and fresh-instance validation

This follow-up is separate from the original frozen inventory. A train-only centered ridge regression with fixed regularization 10 supplies $b+w^\top x$. The original DeepFM fitting procedure learns residual labels, with the same training/validation split and early-stopping rule. Its fitted prediction is added to the ridge backbone; the residual is retained only if it improves the fixed validation MSE. This is a conventional predictor repair, not a newly claimed ridge-learning method. Linear terms are folded into the accessible QUBO and the neural branch retains its stated form. Both full and conditional seeding receive this model.

The first 20 runs reuse the two original 512-bit conditions and all five seeds. They are development evidence: Trinity improves the higher-order mean from 0.62510 to 0.58080, with five paired wins, while ridge–DeepFM A2 reaches 0.57520. On the quadratic condition Trinity has two wins, one tie and two losses; its mean changes from 0.46301 to 0.46156. Table 16 includes every original completed comparator and both modified variants. The lower mean than original BODi on the higher-order development cohort is not independently replicated against BODi.

Table 16: Development comparison on reused original 512-bit instances. Five seeds, 256 true calls; mean $\pm$ sample SD. Original thirteen-method results are retained alongside both predictor predictor configurations. Lower is better; bold marks the lowest displayed mean, not statistical significance.

Method	$\nu = 0$	$\nu = 0.5$
FMQA	0.45391 ± 0.01015	0.60553 ± 0.01196
FM+anc.	0.45258 ± 0.00557	0.60732 ± 0.01273
FM + PROTES	0.46258 ± 0.00316	0.66295 ± 0.00579
A1	0.46437 ± 0.00544	0.66164 ± 0.00647
A2	0.46320 ± 0.00630	0.60840 ± 0.01663
A3	0.46543 ± 0.00562	0.64533 ± 0.00859
A2+	0.45918 ± 0.00360	0.60881 ± 0.01509
Trinity-D	0.46301 ± 0.00642	0.62510 ± 0.01034
Random	0.46555 ± 0.00579	0.66268 ± 0.00655
CEM	0.46219 ± 0.00765	0.63986 ± 0.00875
GA	0.45578 ± 0.00285	0.64941 ± 0.00244
PROTES	0.46461 ± 0.00487	0.66314 ± 0.00573
BODi	$\mathbf{0.42273 \pm 0.01176}$	0.59256 ± 0.00471
A2 (ridge)	0.46590 ± 0.00615	$\mathbf{0.57520 \pm 0.02520}$
Trinity	0.46156 ± 0.00325	0.58080 ± 0.02039

The unchanged predictor configuration was then frozen for seeds 1701–1705 and independently generated instances 17001–17005, at both $\nu = 0$ and 0.5. All 80 planned runs completed. Primary contrasts compare Trinity with Trinity-D within each condition; other comparisons are secondary. All eight methods share exactly the same first 128 designs and labels and receive 256 true calls. New queries are unique, journal counts match checkpoints, and all recorded losses were recomputed from the objective specification. Original and modified sources have separate fixed hashes. There is

no reuse of these new outcomes for parameter selection. This addresses fresh-instance sensitivity in one already selected synthetic family, not broad task selection or optical generalization. Five pairs remain exploratory.

Table 17: Complete fresh-instance check: mean ± sample SD for every planned method, same conditions and budgets as Table 3. The model predictor configuration is applied equally to ridge–DeepFM A2 and Trinity.

Method	$\nu = 0$	$\nu = 0.5$
FM+anc.	$\mathbf{0.45223 \pm 0.01360}$	0.59449 ± 0.02761
PROTES	0.46855 ± 0.00371	0.66340 ± 0.00489
CEM	0.46297 ± 0.00631	0.64264 ± 0.00505
GA	0.45590 ± 0.00518	0.64463 ± 0.00518
A1	0.46348 ± 0.00300	0.66314 ± 0.00603
T-D	0.46211 ± 0.00644	0.63910 ± 0.01255
A2 (ridge)	0.46996 ± 0.00211	$\mathbf{0.56234 \pm 0.03020}$
Trinity	0.46582 ± 0.00807	0.56531 ± 0.01761

These studies add 25,600 valid optimization calls on inexpensive synthetic objectives. An initial launcher imported the original predictor instead of the variant; those attempts were excluded by source identity, retained with their call journals, and rerun without changing the fixed model specification. They were not used to choose settings. The supplemental follow-up archive preserves both protocols, source snapshots, every valid run and the excluded launcher records. Original physical results and original runtime estimates are not relabeled as results of the ridge–DeepFM model. Concurrent execution precludes causal cross-method wall-time claims.

Table 18: Original-study mean full-model scores N_g per 256-call run, five complete repetitions. A2+ and Trinity-D each use the displayed bit-proposal count. These measurements use the original DeepFM predictor.

Condition	A1 N_g	A2+ N_g	T-D N_g	Bit proposals*
WDM 1,024	2,048.0	2,304.0	5,719.6	2,097,152
Polynomial 5,000	2,048.0	2,304.0	22,063.6	10,240,000
Polynomial 10,000	2,048.0	2,304.0	40,769.6	20,480,000

D Development and diagnostic evidence

D.1 Development outcomes and rejected hypotheses

Development is separate from final confirmation. In one 512-variable, 192-call comparison with 64 initial points, A3+ at block size 16 achieved loss 0.65724 on its optimization grid, whereas block size 64 achieved 0.78760, the full-block version 0.80028, GA 0.72341, and anchored FM 0.74737. With fixed physical monitors in subsequent refinement, the selected 16-bit design scored 0.65725 at 20 nm and 0.69047 at 10 nm, while GA scored 0.72193 and 0.69331. The reduced margin is retained rather than reporting only the favorable coarse value.

A separate 256-call condition with 128 initial observations gave A3+ block size 64 a terminal 20 nm loss near 0.62311, compared with GA 0.72267, corrected full seeding 0.78598 and anchored FM 0.74792. Fixed-monitor refinement gave A3+ $0.62260/0.60210$ and GA $0.72273/0.72774$ at $20/10$ nm. Changing both seed and budget prevents treating the difference between these two studies as an isolated effect of initial label count.

The TT learning-rate assay compared 0.01 with 0.0001 on a separate development seed. A3+ reached 0.643265 with the reference rate versus the initial incumbent 0.734020 at the smaller rate; A2+ reached 0.723790 versus 0.734020. Better exact-seed retention alone did not yield better optimization, so the frozen default uses 0.01.

A structured-start development condition at 128 total calls and 64 initial labels reached 0.518919 for A3+ with 64-bit blocks, 0.532446 for A2+, and 0.525126 for BODi. This is one seed, with no fine-grid confirmation of that small BODi margin. Some comparisons ran on different devices, so these development timings do not establish algorithmic wall-clock superiority. The outputs were opened after primary optical settings were frozen and did not change them.

Auxiliary quadratic-branch supervision was also considered. With $e = g - f$,

$$e^2 + \lambda(q - f)^2 = (1 + \lambda)\left(e - \frac{\lambda h}{1 + \lambda}\right)^2 + \frac{\lambda}{1 + \lambda}h^2. \tag{29}$$

This controls cancellation on the same empirical distribution but supplies no off-support prediction guarantee. Its development evidence was mixed, so the primary model keeps its coefficient zero. Similarly, the 72-cell solver assay found exact minima despite an active temperature floor in most cells; it did not justify a speculative annealer modification.

D.2 Two nested oracles: prediction, initialization and observed gain

The outer oracle f produces charged observations; the fixed fitted model g supplies input–output scores to inner PROTES search. Their errors and evaluation counts are distinct. Figure 12 is a post hoc diagnostic from saved checkpoints, with no new simulator calls. Its fixed cohorts are ten film repetitions and five instances at each of 5,000 and 10,000 variables, with every prespecified A3+ run complete. Later completed studies are not added to this illustration.

The median first-fit validation R^2 is approximately -0.0136 for film and -2.1755 for the large controlled study. These are held-out splits of the initial dataset, not uniform error certificates. The median within-run improvement relative to the best initial observation is 16.6% and 0.287%, respectively; the objectives differ, so these percentages do not rank difficulty across applications. Four of ten film runs fail to improve their initial best, whereas all ten corrected-decomposition scaling runs improve it slightly. These findings distinguish useful surrogate prediction from local optimization accuracy and do not attribute causality from a scatter plot.

The TT creator initially places substantial mass on its complete seeds: the median exact-seed mass over outer rounds is about 0.326 for film and 0.331 for scaling, with a median of 21 seed hits in the first 64 draws in both cohorts. Subsequent PROTES updates can move this mass elsewhere. A small final mass on the exact seed vectors is not itself failure: the search may explore useful neighbors. Consequently, neither initialization likelihood nor its subsequent decrease establishes true-objective improvement. This distinction between generative-model fidelity and optimization performance also appears in tensor-network estimation-of-distribution algorithms (Gardiner & Lopez-Piqueres, 2026). The recorded trace supports separate accounting of seed transfer and outer-loop performance.

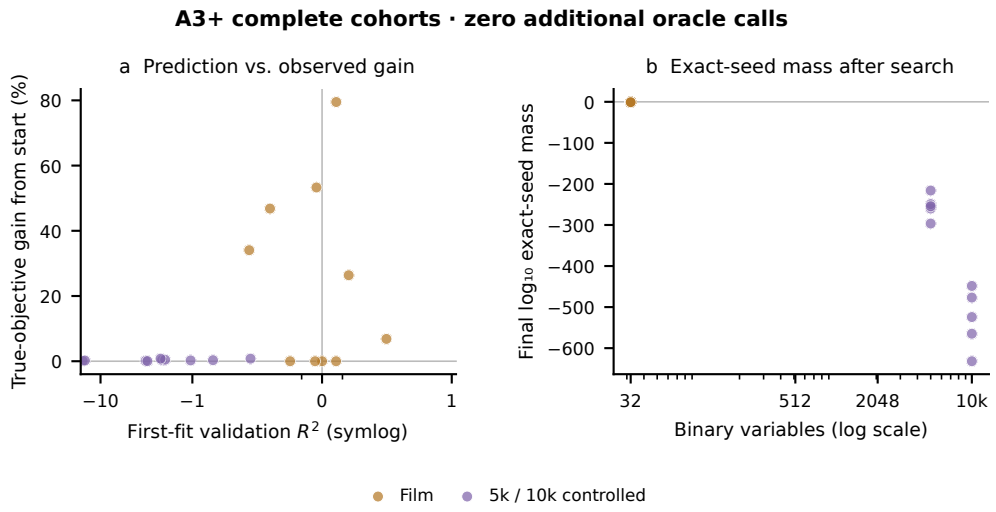

Figure 12: Completed-cohort A3+ diagnostics. Left: each point pairs first-fit validation R^2 with $100(f_{\text{initial}} - f_{\text{terminal}})/|f_{\text{initial}}|$; the horizontal axis uses a symmetric logarithmic scale with a linear region around zero. Right: each point is a run's median final log-probability mass on its exact seed set across outer rounds. Initial seed masses are near one third. All observations are retained; no jitter, smoothing, causal claim or extra physical evaluation is used.

E Quantum interface and runtime accounting

Table 19: Measured Trinity costs: mean seconds over five completed A3+ runs per condition, each with 256 true calls. N_Q counts QUBO subproblems. TT combines initialization and full-surrogate search; Other includes correction, context handling, genetics, bookkeeping and remaining host work. These are recorded execution costs, not causal cross-host comparisons. Physical refinement and visualization are excluded; no QPU was used.

Condition	N_Q	Total	True f	Fit g	Ising	TT	Other
WDM 1,024	512	436.48	410.35	2.50	8.11	13.19	2.34
Polynomial 5,000	2528	36.44	0.05	4.82	15.99	6.21	9.38
Polynomial 10,000	5024	84.71	0.09	8.37	31.22	8.45	36.58

A backend receives conditional QUBO or exactly converted Ising coefficients and returns logical assignments and metadata. Logical variable count is distinct from physical embedding resources. Batching and embedding reuse are possible engineering choices, not measured performance here. A serial accounting model is

$$T_C = T_{\text{other}} + N_{\text{sub}} t_C, \tag{30}$$

$$T_Q = T_{\text{other}} + T_{\text{embed}} + N_{\text{sub}} t_Q + T_{\text{queue}}, \tag{31}$$

$$t_Q = t_{\text{program}} + t_{\text{access overhead}} + t_{\text{network}} + R_{\text{reads}}(t_{\text{anneal}} + t_{\text{readout}} + t_{\text{delay}}). \tag{32}$$

The 1 ms scenario is not a current device specification. D-Wave's documented timing example reports a 20 μs anneal, 39.76 μs readout and 21.02 μs delay, but 15.9354 ms programming and 10.42682 ms additional access overhead. Its charged access time is 16.01618 ms; adding that overhead gives 26.443 ms before network and queue costs (D-Wave Quantum, 2026b). This illustrative two-spin call is not a measurement of our embedded subproblems. Actual costs depend on the processor, embedding and read count; the documented programming time alone is typically 10–20 ms. Reusing an embedding does not eliminate coefficient programming for different QUBOs. A serial current-device deployment therefore cannot infer a 1 ms all-in cost from the anneal duration. Future amortization or parallel packing would require separate capacity, quality and timing evidence.

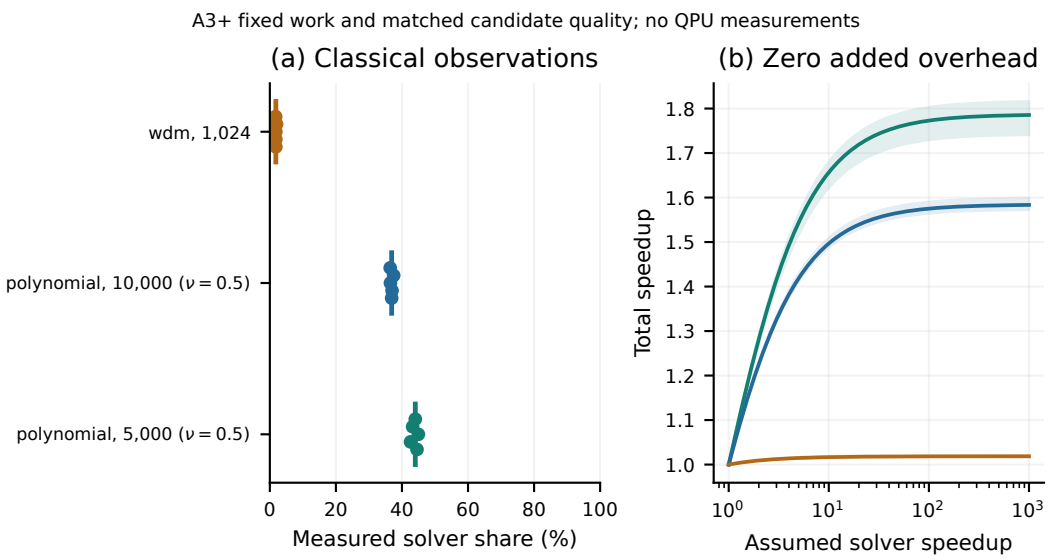

Figure 13: Observed A3+ classical solver fractions for completed scaling and 1,024-bit optical runs, with hypothetical fixed-work replacement. Each point is one completed run; bands on the right are min/max over runs, not confidence intervals. Added overhead is set to zero and useful-candidate quality is assumed matched. No QPU measurements enter either panel. Separate physical refinement and figure-generation work are excluded.

For example, the hypothetical 10^4-subproblem workload above yields the following sensitivity to work outside the solver, with all times in seconds:

Other work	Classical total	Quantum total	Assumed speedup
0	10000	10	1000×
100	10100	110	91.8×
1000	11000	1010	10.89×
10000	20000	10010	2.00×

This table assumes classical time 1 s and quantum total per-subproblem time 1 ms at matched useful-candidate quality, with no additional setup or queue delay. The assumed classical time is much larger than the approximately 6 ms per-subproblem time in our current scaling implementation. It describes a potential future workload, not a re-interpretation of current measurements. A quantum mechanism for attaining that quality and cost requires hardware evidence. Additional trials can translate speed into better solutions only through their success probabilities, which may differ across backends.

Overlapping stages requires a separately measured model. T_{other} includes the true oracle and every host-side search stage. Even with instantaneous quantum computation, mean amortized overhead per subproblem must be below measured $T_{\text{solver}}/N_{\text{sub}}$ for any net speedup. The completed large-study runs give approximately 5.7–7.7 ms for this threshold. It is not a claim that a QPU meets it.

Applying the same assumed all-in cost of 1 ms to the actual recorded workload gives $T_Q^{\text{hyp}} = T_C - T_{\text{solver}} + 0.001 N_{\text{sub}}$. The 5,000-bit A3+ runs make 2,528 calls each, yielding hypothetical total speedups 1.550–1.621 (median 1.591); the 10,000-bit runs make 5,024 calls each, yielding 1.439–1.461 (median 1.448). Thus the proposed interface has a quantitative potential benefit under this cost and quality assumption even though the current classical subproblems take milliseconds rather than the one second assumed in the earlier illustrative scenario. Global setup/queue costs are still set to zero, and no quantum observation is implied.

Amortizing overhead across independent subproblems. Ocean's parallel-embedding interface supports different BQMs on disjoint embeddings (D-Wave Quantum, 2026a). A proposed batched

implementation may group product/anchor-context calls within a fixed controller pass, but must preserve the update barriers between passes and outer rounds. If pass j contains K_j calls and the hardware can accommodate P disjoint embeddings, its batch count is $B_P = \sum_j \lceil K_j/P \rceil$. With unchanged host work and matched useful-candidate quality, it improves on the recorded serial solver only if

$$T_{\text{setup}} + B_P\, t_{\text{batch}} < T_{\text{solver}}. \tag{33}$$

The recorded scaling workload has 32 passes, with 79 or 157 calls per pass. The following strict upper thresholds for all-in batch time (ms) use zero added setup cost; ranges span the five recorded runs. They measure allowable cost, not predicted QPU latency.

Hypothetical slots P	5,000 bits	10,000 bits
1	5.9–7.7	5.7–6.9
4	23.3–30.5	22.5–27.2
8	46.6–60.9	45.1–54.4
16	93.1–121.9	90.1–108.8

Slot counts are hypothetical: physical embedding capacity and quality have not been verified, and readout cost can increase with the occupied chip area. The comparator is the measured serial classical solver; establishing a hardware advantage also requires a fair parallel classical comparison. Thus packing supplies a specific engineering hypothesis for amortizing programming overhead, not a quantum speedup result.

Binary coefficient errors $\Delta a_i, \Delta B_{ij}$ give energy-error oscillation at most $\zeta = \sum_i |\Delta a_i| + \sum_{i<j} |\Delta B_{ij}|$. A $\widetilde{\eta}$-optimal distorted-energy candidate is consequently $(\widetilde{\eta} + \zeta)$-optimal for the intended quadratic; constants cancel. Ising features in $\{-1, 1\}$ give a factor-two bound. Chain repairs require their own error accounting, and matched energy need not imply matched candidate diversity.