NEW RECORDS FOR THE HADAMARD MAXIMAL DETERMINANT PROBLEM

Anonymous authors
Paper under double-blind review

ABSTRACT

The Hadamard Maximal Determinant problem seeks an $n \times n$ matrix X with entries in $\{\pm 1\}$, which maximizes the determinant $|\det X|$ for a given order $n \in \mathbb{N}$. We report matrices attaining new record determinants for orders $n \equiv 3 \pmod 4$ between 103 and 119, as well as $n = 51$. We provide a proof of optimality within a specific family of circulant–block constructions for X. For these orders, we recast the Hadamard problem as a sequence reconstruction problem from a pair of periodic autocorrelation functions, subject to certain arithmetic conditions. We describe several learning–based strategies for reconstruction and provide a comparison with classical simulated annealing–style approaches as a benchmark.

1 INTRODUCTION

The *Hadamard maximal determinant problem*, originally posed by Hadamard (1893) is to find an $n \times n$ $\{\pm 1\}$–matrix with maximum possible determinant for a given order n:

$$\mathcal{D}(n) = \max \left\{ |\det X| \mid X \in \{-1, +1\}^{n \times n} \right\} .$$

Naïvely, the search space for the problem grows superexponentially as 2^{n^2}. The problem of finding a general construction for all orders n has remained unsolved for over a hundred years. One reason is that the problem naturally divides based on the residue class of n (mod 4). The most progress has been made for $n \equiv 0$ (mod 4) (Proposition 4, Browne et al. (2021)), and the least progress for $n \equiv 3$ (mod 4). Certifiably maximal determinant matrices have been found for all orders up to $n = 22$ inclusive Orrick (2012), with the $n = 23$ record conjectured to be maximal. Various constructions have been made in the literature for special infinite families of solutions, and there exist certain ad–hoc examples, but no general construction has been found for any residue class.

Geometrically, the determinant of the matrix X_n gives the signed volume of an n–dimensional parallelotope in $\mathbb{R}^n$ spanned by the rows/columns of X_n, whose vertices lie on the corners of the hypercube with edge length 2. The volume of X_n is maximized when the defining vectors are orthogonal, yielding an elementary upper bound for all orders as $|\det X_n| \leq n^{n/2}$. This bound is sharp only when the orthogonality condition is met, which, for $n > 2$, is only possible when $n \equiv 0$ (mod 4). For other orders, one must find an optimal relaxation of the orthogonality condition, with different arithmetic 'defects' to orthogonality arising for each order $n \not\equiv 0$ (mod 4).

The associated Gram matrix $G_n = X_n X_n^\top$ encodes the geometry of the row vectors of X_n: its diagonal entries are their squared lengths, while its off-diagonal entries are their pairwise inner products. Choosing a suitable normalization, it is not hard to show that the off-diagonal elements of the Gram matrices corresponding to the maximal determinant matrices assume values $\equiv n$ (mod 4) Brenner (1972). For $n \equiv 1$ (mod 4), the natural symmetric ansatz $G_n = (n - t)\mathbf{1}_n + tJ_n$, where J_n is the uniform $n \times n$ matrix of 1s, yields an upper bound on $\det X_n$ which is sharp for at least finitely many matrices of such order. A similar ansatz exists for $n \equiv 2$ (mod 4) Browne et al. (2021). This symmetry breaks in $n \equiv 3$ (mod 4) — neither of the uniform choices $t = -1, 3$ are satisfactory, and the matrices which yield natural upper bounds arrange themselves into a block–diagonal structure over a uniform (-1)–background Ehlich (1964a;b). We will provide new record–achieving constructions for maximal determinant matrices in orders $n \equiv 3$ (mod 4) that may be realised as *circulant matrices*. In this case, basic Fourier theory states that the problem of constructing the maximal determinant matrices is equivalent to a sequence reconstruction problem from analytically computed target data.

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

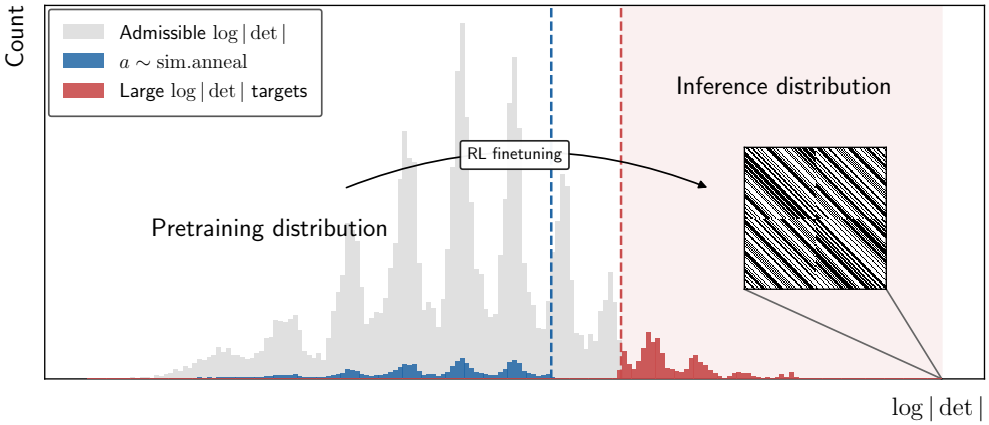

Figure 1: Maximal determinant problem reformulated as PAF reconstruction problem using RL.

The problem serves as an excellent benchmark for testing the capabilities of reinforcement-learning approaches in extremely sparse-reward environments. Recent successes, as evidenced by the many applications of RL to combinatorial optimization Kool et al. (2019); Kwon et al. (2020); Chalumeau et al. (2023); Kim et al. (2024), motivate us to apply similar techniques to search for maximal determinant targets which improve currently known lower bounds Browne et al. (2021); Orrick (2012).

2 Periodic Autocorrelations

Here we set up the framework for reformulating the maximization of the determinant as a sequence reconstruction problem Kotsireas et al. (2019). In this work, we focus on $\{\pm 1\}$-valued $n \times n$ matrices X constructed via $m \times m$ circulant matrices A and B, where $m = (n-1)/2$. We consider two main constructions, *uncorrelated* $X(A, B)$ and *correlated* $X^{\text{corr.}}(A, B)$, defined in the following manner:

$$X(A,B) := \begin{pmatrix} 1 & j^T & -j^T \\ j & A & B \\ -j & B^T & -A^T \end{pmatrix}, \quad X^{\text{corr.}}(A,B) := \begin{pmatrix} 1 & j^T & -j^T \\ j & A & B \\ j & -B & -A^T \end{pmatrix} \tag{1}$$

where j is an all-ones vector of size m. For any fixed order $n \equiv 3 \pmod 4$, we will therefore focus on finding the circulant matrices A and B which maximize the determinant $|\det X(A,B)|$ (or $|\det X^{\text{corr.}}(A,B)|$, whichever is larger). We shall focus on both cases separately.

An $m \times m$ circulant matrix A is uniquely described by a sequence $\{a_j\}_{j=0}^{m-1}$ via $A_{ij} = a_{j-i \pmod m}$. For a circulant matrix A constructed in such way from a sequence a, we shall use notation $A = \text{circ}(a)$, similar to the one used in Brent & Yedidia (2018).

Uncorrelated X: From the construction of putative/candidate Gram matrices in Brent et al. (2011); Ehlich (1964b); Orrick (2005), one key desirable property of the Gram matrix G of X is the condition $G_{ij} \in \{-1, +3\}$ for all $i \neq j$. Let $X = X(A,B)$, then the Gram matrix G of X is given by:

$$G = XX^T = \begin{pmatrix} n & \gamma_0^T & \gamma_1^T \\ \gamma_0 & C & -j \otimes j \\ \gamma_1 & -j \otimes j & C \end{pmatrix}, \tag{2}$$

where $\gamma_0 = j + (A-B)j$, $\gamma_1 = -j + (B^T + A^T)j$, and $C = AA^T + BB^T + j \otimes j$. Since A and B are circulant, we may describe them using sequences $\{a_j\}_{j=0}^{m-1}$ and $\{b_j\}_{j=0}^{m-1}$, respectively. This allows us to identify border terms γ_i with segment sums as $\gamma_0 = (1 + |a| - |b|)j$, $\gamma_1 = (-1 + |a| + |b|)j$, where $|a| := \sum_j a_j$ and $|b| := \sum_j b_j$ are segment sums.

Furthermore, the diagonal component C is also a circulant matrix:

$$C = \mathrm{circ}(\{\mathrm{PAF}_a(k) + \mathrm{PAF}_b(k)\}_{k=0}^{m-1}) + j \otimes j \tag{3}$$

where PAF_a and PAF_b are the periodic autocorrelation functions of sequences a, and b, defined below.

Definition 2.1. Let $\{a_j\}_{j=0}^{m-1}$ be a sequence of integers. The *periodic autocorrelation function* $\mathrm{PAF}_a(s)$ is defined as:

$$\mathrm{PAF}_a(s) = \sum_{j=0}^{m-1} a_j a_{j+s \pmod m}\,, \tag{4}$$

defined over all $0 \le s < m$.

The Ehlich Gram constraint $G_{ij} \in \{-1, +3\}$ for all $i \neq j$ imposes the following condition on the sum of the periodic autocorrelation functions:

$$\forall s > 0 \colon p(s) := \mathrm{PAF}_a(s) + \mathrm{PAF}_b(s) \in \{-2, +2\}\,, \tag{5}$$

where $p(0) = 2m = n - 1$. We will refer to such functions as *admissible*. As implied by the proof of Proposition 3 of Armario & Flannery (2020), a pair a, b of binary sequences whose periodic autocorrelation functions satisfy (5) corresponds to a pair of subsets $B, D \subseteq \mathbb{Z}_m$ forming a $2 - \{m; |B|, |D|; \mu\}$ almost supplementary difference set, where $\mu = |B| + |D| - \frac{m+1}{2}$. Notably, Theorem 3 of the same work identifies this same value of μ in the characterization of almost supplementary difference sets associated with quaternary sequences having optimal autocorrelation.

By summing the function $p(s)$ over $0 \le s < m$, we obtain well-known Diophantine equation constraint on the segment sums Djokovic & Kotsireas (2011); Williamson et al. (2026):

Proposition 2.2. *Let N be the number of times $p(s) = +2$ for $s \neq 0$, then:*

$$|a|^2 + |b|^2 = 4N + 2\,. \tag{6}$$

Proof. The proof of the original statement can be found in Djokovic & Kotsireas (2011). We provide a proof for our parametrization in Appendix C. □

This provides us with the first-stage filtering, where we identify pairs of segment sums $(|a|, |b|)$ satisfying (6). The second-stage filtering is based on identifying the determinant associated to an admissible PAF-sum function $p(s)$ and enumerating the ones which exceed the current record Orrick (2012).

We proceed by decomposing X into its Fourier modes. In particular, let A be a circulant matrix, then its normalized eigenvectors are given by:

$$v_t = \frac{1}{\sqrt{m}}(1,\ \zeta_m^t,\ \ldots, \zeta_m^{(m-1)t})^T, \quad 0 \le t < m\,, \tag{7}$$

where $\zeta_m := \exp(2\pi i/m)$ is the primitive m-th root of unity. It follows that the eigenvalues are $\lambda_t = \sum_{k=0}^{m-1} a_k \zeta_m^{kt}$. Let $F = (v_0, \ldots, v_{m-1})$, then F diagonalizes A into Fourier modes:

$$F^\dagger A F = \mathrm{diag}(\widehat{a}_0, \ldots, \widehat{a}_{m-1})\,. \tag{8}$$

By taking a unitary transform of X using $U := 1 \oplus F \oplus F$, we obtain a decomposition:

$$U^\dagger X U \cong \widehat{X}_0 \oplus \bigoplus_{t>0}^{m-1} \widehat{X}_t, \quad \text{(up to permutations)} \tag{9}$$

where we reorder the basis to group the $t = 0$ modes together, and the respective $t > 0$ modes for A and B together. Note that j only couples to the zero-modes, so $\widehat{X}_0$ is the 3×3 matrix corresponding to the zero-mode of X incorporating the border terms:

$$\widehat{X}_0 = \begin{pmatrix} 1 & \sqrt{m} & -\sqrt{m} \\ \sqrt{m} & \widehat{a}_0 & \widehat{b}_0 \\ -\sqrt{m} & \widehat{b}_0 & -\widehat{a}_0 \end{pmatrix}\,, \tag{10}$$

whereas the higher modes $\widehat{X}_{t>0}$ are given by:

$$\widehat{X}_{t>0} = \begin{pmatrix} \widehat{a}_t & \widehat{b}_t \\ \overline{\widehat{b}}_t & -\overline{\widehat{a}}_t \end{pmatrix} . \tag{11}$$

Since U is unitary, it preserves determinant, therefore, we obtain:

$$\det X = \det X_0 \prod_{t>0}^{m-1} \left(|\widehat{a}_t|^2 + |\widehat{b}_t|^2 \right) . \tag{12}$$

The zero-mode determinant is straightforward (it depends only on zero modes of A and B), however, the higher modes may be related to the Fourier transforms of the respective PAFs in the following manner: using the Wiener-Khinchin theorem, the power spectrum of a sequence $(a_0, \ldots, a_{m-1})$ is the DFT of its periodic autocorrelation function:

$$|\widehat{a}_t|^2 = \sum_{s=0}^{m-1} \mathrm{PAF}_a(s) \zeta_m^{st} , \tag{13}$$

therefore, the sum of magnitudes $|\widehat{a}_t|^2 + |\widehat{b}_t|^2$ may be rewritten purely in terms of the Fourier transform of the PAF-sum function $p(s)$ as: $\widehat{p}_t = |\widehat{a}_t|^2 + |\widehat{b}_t|^2$. The Fourier transform $\widehat{p}_t$ is given by:

$$\widehat{p}_t = 2m + 2 \sum_{s=1}^{(m-1)/2} p(s) \cos\left(\frac{2\pi st}{m} \right) , \tag{14}$$

therefore, in order to evaluate the determinant of a proposed PAF-sum target, we may do so directly using its values, without requiring its reconstruction into sequences a and b.

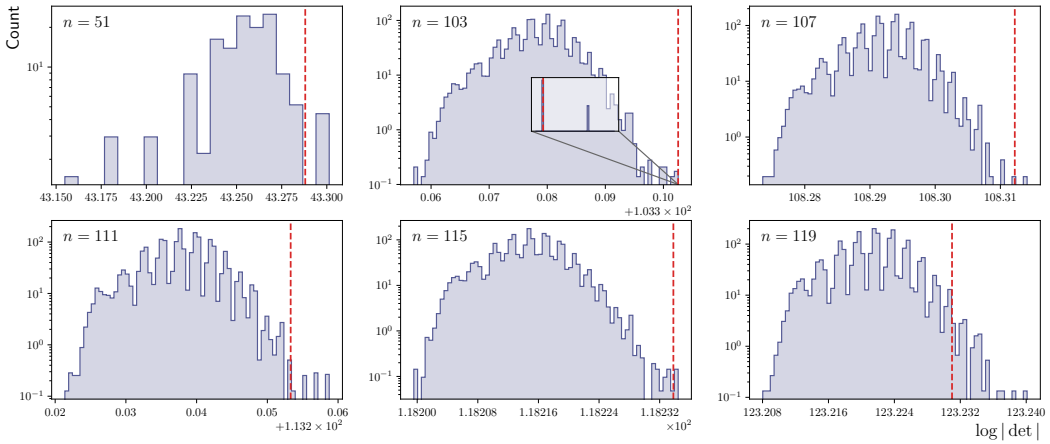

Figure 2: Distribution of $\log|\det|$ values of all admissible PAF-sum targets computed using (12). The dashed red line (----) denotes the current record in Orrick (2012).

Once we have a list of candidate PAF-sums $\{p_i(s)\}_i$ achieving maximal possible determinant in the uncorrelated ansatz class $X = X(A, B)$ (1), we want to reconstruct the original sequence pair (a, b). This is an undetermined system of quadratic equations in $2m$ variables on the product of boolean cubes $\{\pm 1\}^m \times \{\pm 1\}^m$. Equivalenty, we would like to minimize the following objective:

$$\min_{a,b \in \{\pm 1\}^m} H(a, b) = \min_{a,b \in \{\pm 1\}^m} \sum_{s=1}^{(m-1)/2} \left| \mathrm{PAF}_a(s) + \mathrm{PAF}_b(s) - p(s) \right|^2 , \tag{15}$$

where $s = 0$ is omitted as it gives constant contribution and sum goes until $(m-1)/2$ as $p(s)$ is reflection-symmetric about the midpoint.

In order to construct a distribution of all PAF-sum targets w.r.t. their expected determinants, we first exhaustively search through the solutions of the Diophantine equation (6) and for each candidate

segment sum pair $(|a|, |b|)$, we sweep through all possible PAF-sum targets. Given that a specified segment sum pair specifies the number of spots where $p(s) = 2$ at $s \neq 0$ and noting that $p(s) = p(-s)$, it is immediately obvious that the total number of all possible PAF targets is:

$$\#\text{admissible targets with segment sum pair } (|a|, |b|) = \begin{pmatrix} \frac{m-1}{2} \\ \frac{|a|^2 + |b|^2 - 2}{8} \end{pmatrix}, \tag{16}$$

where the dimension which has the largest number of admissible targets is $n = 119$, which has 475020 admissible targets, which is small enough to be enumerated completely. Thus reconstructing a maximal PAF-sum target immediately implies a verifiable certificate of maximality within the family $X(A, B)$ with $p(s \neq 0) \in \{-2, 2\}$. The distribution of all such targets is shown in Fig. 2. Correlated $X^{\text{corr.}}$ has a similar construction, described in Appendix C.1.

3 Maximal determinants as sequence reconstruction

Finding the maximal determinant representatives of the families (1) may be regarded as a reconstruction problem which attempts to construct sequences $(\{a_j\}_{j=0}^{m-1}, \{b_j\}_{j=0}^{m-1})$ satisfying the relevant constraints depending on whether they parametrize $X(A, B)$ or $X^{\text{corr.}}(A, B)$. While $X(A, B)$ matrices allow every pair of sequences with values in $\{\pm 1\}$ minimizing the energy objective (15), $X^{\text{corr.}}(A, B)$ additionally requires two more conditions (See appendix C.1):

$$|a||b| = m, \quad a * c = 2, \quad \text{where } c_t = b_t + b_{-t}\,. \tag{17}$$

In both cases, we define the state space $\mathcal{S}(|a|, |b|, m)$ as the set of all pairs of sequences (a, b) of length m with segment sums $\sum_j a_j = |a|$ and $\sum_j b_j = |b|$. The initial state s_0 is sampled uniformly from the state space. Since we are interested in reconstructing PAF-sum targets using sequences with pre-determined segment sums $|a|$ and $|b|$, we define action space $\mathcal{A}(|a|, |b|, m)$ to be the set of all moves which swap two distinct-valued entries within sequence a or sequence b. From this, it immediately follows that the state and action space sizes scale as:

$$|\mathcal{S}(|a|, |b|, m)| = \begin{pmatrix} m \\ \frac{m + |a|}{2} \end{pmatrix} \begin{pmatrix} m \\ \frac{m + |b|}{2} \end{pmatrix}, \quad |\mathcal{A}(|a|, |b|, m)| = \frac{2m^2 - |a|^2 - |b|^2}{4}\,. \tag{18}$$

Typical scales of state spaces that we consider are shown in Table 5. Due to the extreme combinatorial explosion, the performance of our methods depends significantly on the design choices for learning useful representations and the reward signals.

3.1 Arithmetic and hierarchical reconstructions

At the orders we are interested in, as indicated in Table 5, the search space is prohibitively large for any direct reconstruction algorithm relying on either reinforcement-learning methods or classical techniques. We therefore break down the reconstruction problem into two steps:

1. Given a PAF-sum target $p(s)$, find the space of all feasible sequences $\{a_j\}_{j=0}^{m-1}$ such that the residual:
$$\text{PAF}_b(s) = r(s) := p(s) - \text{PAF}_a(s) \tag{19}$$
is algebraically unobstructed.
2. Given an algebraically unobstructed residual $r(s)$, attempt to perform an exhaustive search for a sequence $\{b_j\}_{j=0}^{m-1}$ solving (19).

Motivated by the use of cyclotomic fields in the recent works Djokovic & Kotsireas (2011); Kotsireas et al. (2025), we also consider a projection of a sequence $\{a_j\}_{j=0}^{m-1}$ to a cyclotomic field $K_d := \mathbb{Q}(\zeta_d)$ for $d \mid m$, and rewrite the residual equation (19) as a norm equation:

$$b_d b_d^* = r_d\,. \tag{20}$$

From this identification, it follows that all obstructions to solvability of the field-norm equation may be detected by studying the class of r_d in Galois cohomology $H^2(K_d/K_d^+)$ (See Appendix E for more details).

3.2 RECONSTRUCTIONS THROUGH LIFTS

Let $m = (n-1)/2$ be a non-prime odd number and $d \mid m$ be one of the factors such that $d > 2$ and $d \neq m$. While the arithmetic handles obstructions at the level of $\mathbb{Q}(\zeta_d)$, we may also consider projections of form:

$$\pi_d \colon \mathbb{Z}[C_m] \to \mathbb{Z}[C_d]\,, \tag{21}$$

where we first attempt to construct a set of potential solutions in $\mathbb{Z}[C_d]$ and then lift them to the full solution in $\mathbb{Z}[C_m]$. More specifically we run two exhaustive searches:

- The coefficients of $\pi_d(b)$ have values in $\{-k, -k+2, \ldots, k\}$ where $k = m/d$. Projections preserve segment sums and map PAFs to PAFs. Therefore, we exhaustively construct a set:
$$\mathcal{P}_d(r, |b|) = \{g \in \{-k, -k+2, \ldots, k\} \mid |g| = |b|,\ \mathrm{PAF}_g = \pi_d(r)\}\ . \tag{22}$$
- If $\mathcal{P}_d(r, |b|) \neq \emptyset$, we search for all possible $\{\pm 1\}$ sequences with prescribed fiberwise sums from $g \in \mathcal{P}_d(r, |b|)$:
$$\mathcal{B}_d(g) = \left\{ b \in \{\pm 1\}^m \;\middle|\; \forall u \in \mathbb{Z}_d \colon \sum_{j=0}^{k-1} b_{u+jd} = g_u \right\}\ . \tag{23}$$
The elements of $\mathcal{B}_d(g)$ are the candidate reconstructions of the PAF equation. We then exhaustively check $\mathrm{PAF}_b = r$ for every $b \in \mathcal{B}_d(g)$.

We massively parallelize the checks (22) and (23) over our custom CUDA kernels. This algorithm defines an **oracle** B that given a candidate sequence $A \in \{\pm 1\}^m$ and a target joint autocorrelation function $p(t) \in \{\pm 2\}^m$, $t > 0$, is able to provide a valid partner B satisfying the conjectured maximal–determinant condition (5), or certify that no such partner exists, i.e. $\mathsf{B}(p; A) := \mathbf{1}\,[\exists\, B \in \{\pm 1\}^m : \mathrm{PAF}_A + \mathrm{PAF}_B = p]$. The problem then reduces to construction of A, and the effective horizon length halves.

3.3 ARITHMETIC RL

The major difficulty of two-stage reconstruction is finding a large number of sequences $\{a_j\}_{j=0}^{m-1}$ which pass the field norm obstructions (43). Although passing these obstructions provides a relatively sparse signal, the decomposition of the global obstruction problem to local tests over relevant places $0 \neq v \in \mathrm{Spec}\ \mathcal{O}_{K_d^+}$ allows us to construct a potential which provides useful feedback for guiding an agent. Therefore, we build an agent, trained using an algorithm similar to that of our RL setup, described in Sec. D.3, whose main objective is to sample a large number of unique sequences with unobstructed residuals for any prescribed PAF-sum target.

Single-PAF GNN model: In order to ensure equivariance, we represent the sequence $\{a_j\}_{j=0}^{m-1}$ as a complete graph $\mathcal{K}_m$, where to each edge e_{ij}, we assign features:

$$e_{ij} = \begin{pmatrix} p_{d(i,j)} \\ r_{d(i,j)} \end{pmatrix}\ , \tag{24}$$

where $d(i,j) = \min(|i-j|, m-|i-j|)$, while the node features are given by $x_i = \phi_{\mathrm{sign}}(a_i) + \phi_{\mathrm{tgt}}(r,t)$, where ϕ_{sign} and ϕ_{tgt} are encoding layers for node and target features, respectively.

Reward shaping: We are after two key objectives: given a PAF-sum target $p(s)$:

- Find sequence a, such that the Galois cohomology class of the projection of the residual $r_d = p_d - a_d a_d^* \in K_d$ vanishes in $H^2(K_d/K_d^+)$ for all $d \mid m$.
- Maximize the number of unique sequences a sampled using RL.

Based on this, we first define a potential based on the local field norm obstructions from Thm. E.2:

$$V_d(a) = -\frac{1}{[K_d^+ : \mathbb{Q}]} \sum_{\substack{\text{Embedding} \\ \sigma \colon K_d^+ \to \mathbb{R}}} \mathbf{1}(\sigma(r_d) < 0) - \sum_{\substack{v \in \mathrm{Spec}\ \mathcal{O}_{K_d^+} \\ v \neq 0}} \mathbf{1}\left((r_d, \Delta_d)_v \neq 1\right)\ , \tag{25}$$

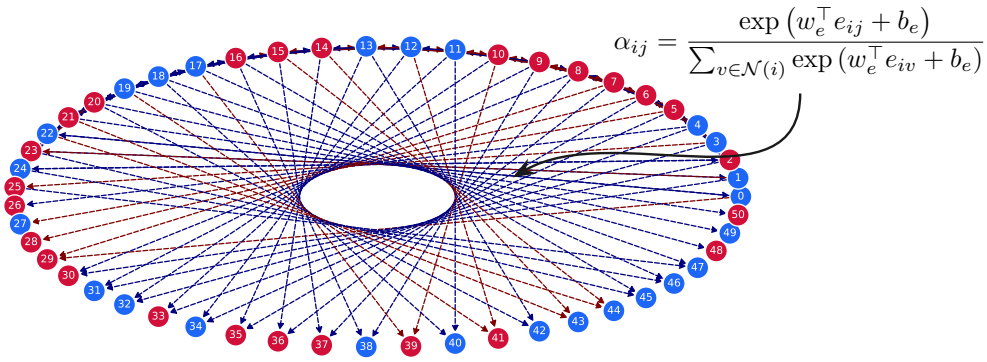

Figure 3: Single-PAF GNN model. Nodes represent individual sequence elements a_j, where red nodes denote $a_j = +1$, while blue nodes denote $a_j = -1$. The edges are colored based on the values of message weights α_{ij}. For each node, only top 2 incident edges with the largest weights are shown.

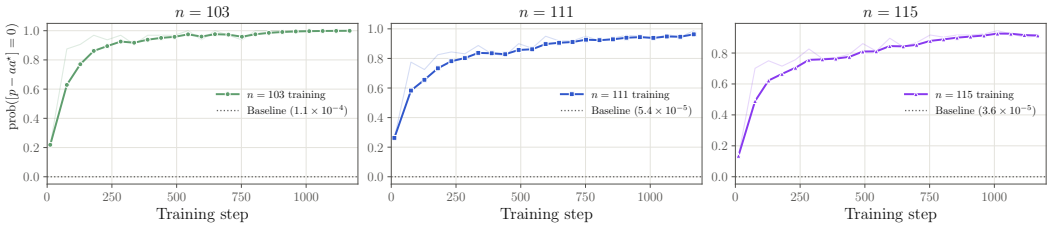

Figure 4: Success rate curves for different orders of the RL agent throughout the training. The success is defined as the environment terminating before truncation due to encountering a state with unobstructed residual relative to maximal determinant PAF-sum target. The dotted horizontal lines denote the baseline probabilities defined as the likelihood of randomly sampled sequence a with $|a| = 1$ having unobstructed residual for the maximal determinant target.

which allows us to shape reward function as:

$$r(s', a, s) = r_{\text{term}}(s', a, s) + r_{\text{novel}}(s', a, s) + \sum_{d|m,\ d>2} \left(\gamma V_d(s') - V_d(s)\right) , \tag{26}$$

where $r_{\text{term}}(s', a, s)$ provides a constant reward of c_{term} if the residual associated to s' is unobstructed, otherwise it's zero, and $r_{\text{novel}}(s', a, s) = \mathbf{1}\left(s' \notin \mathcal{D}_{\text{seen}}(p) \ \wedge\ [p - s's'^*] = 0\right)$ incentivizes previously unobserved sequences a for a given PAF-sum target p, where we maintain a list $\mathcal{D}_{\text{seen}}(p)$ of all sequences that have been observed to have unobstructed residuals w.r.t. target p. In order to further incentivize the novelty of the constructed sequences, we also add additional reward term coming from random network distillation Burda et al. (2019). The training curves of this model are shown in Fig. 4, while quantitative reconstruction results are in Table 1.

3.4 Arithmetic–free baseline

To understand the utility of the obstruction–theoretic machinery constructed above, we develop a method that makes no assumption about the underlying structure of the binary sequence A.

3.4.1 Environment

The core idea is to work with gaps between consecutive $+1$ elements in the sequence A, thereby quotienting out part of the shift symmetry of the autocorrelation function. Denote by $S_A \subseteq \mathbb{Z}_m$ the support of the $+1$ elements in A, and the cardinality $k_A := |S_A| = (\sigma_A + m)/2$. Let $c_i \in S_A$, then our state space consists of a stream

$$\mathcal{S} := \{(g_1, \ldots, g_{k_A - 1}) : g_i = (c_{i+1} - c_i) \in \{1, \ldots, m - k_A + 1\}\} , \tag{27}$$

which is produced autoregressively by the model. The action space consists of a discrete choice of gap g_i at each step along the trajectory. Note the final gap is forced and we rotate the stream s.t. $c_1 = 0$. Suppose we have committed the first t gaps and have a partial reconstruction of A, with the last committed position c_i. Define $R_t = m - c_t$ as the remaining length to be filled at step t and $L = k_A - c_t$ as the remaining length of the trajectory. Then the next gap $g_{t+1} \in \{1, \dots, R_t - L_t + 1\}$, and we mask actions accordingly.

Our rewards can only depend on A since we have no access to B during the rollout. Note that Parseval's theorem says that the Fourier transform pair $(\Delta, \hat{\Delta}_t)$ satisfies

$$\Delta(s) := \mathrm{PAF}_A + \mathrm{PAF}_B - p \,, \quad \sum_s \Delta(s)^2 = \frac{1}{m} \sum_t |\hat{\Delta}_t|^2$$

From the definition of the Hamiltonian, $\sum_t |\hat{\Delta}_t|^2 \propto H$, and hence a necessary condition for A to be a feasible sequence is that $\hat{\Delta}_t$ vanishes for a complete trajectory. Noting $\hat{\Delta}_t = |\hat{a}_t|^2 + |\hat{b}_t|^2 - \hat{p}_t$, we see that non–negativity of $\hat{b}_t$ requires the lower bound

$$\underline{H}(A) := \sum_t \left[\max(0, |\hat{a}_t|^2 - \hat{p}_t)\right]^2 \leq \sum_t |\hat{\Delta}_t|^2 \,, \tag{28}$$

on the Hamiltonian must vanish in order for A to admit a partner. We encourage this behaviour via the following reward, awarded on terminal states,

$$g(A) := \mathbf{1}\left[|\hat{a}_t|^2 \leq \hat{p}_t : t = 1, \dots, h\right] \,. \tag{29}$$

While this is a necessary condition for the existence of a partner to A, the reward only accounts for feasibility of A without any reference to B. Surprisingly, this provides sufficient signal for the trained policy to efficiently sample realizable pairs (A, B). We attempted to augment this with a terminal reward from the oracle on the complete A, $r(s_t) \mapsto r(s_t) + \lambda \cdot \mathbf{1}[t = T] \cdot \mathsf{B}(A(s_t))$, $\lambda \in \mathbb{R}^+$, but the oracle rewards were too sparse to have a significant effect on performance.

3.4.2 Model

To motivate our design, note our strategy for reconstructing the pair of sequences which yield a target joint autocorrelation function $p(t)$ is to sample in parallel from the learned policy π_θ to yield A, then obtain B or a certificate of infeasibility of A from an oracle, which can be constructed for orders n where $m = (n-1)/2$ is not prime. Note that the question of whether a given A has a feasible partner to yield a given target is deterministic, and all the randomness arises from the policy. Define the pass@1 rate as $\rho_1 := \mathbb{P}_{A \sim \pi_\varphi}[\mathsf{B}(p(t); A(s_t)) = 1]$. It will also be useful to record the probability that at least a single feasible example is recovered from k samples, averaged over the validation set V — the pass@k rate, which we estimate as

$$\rho_k = \frac{1}{|V|} \sum_{i \in V} \left(1 - \binom{n_i - c_i}{k} \Big/ \binom{n_i}{k}\right) \tag{30}$$

with c_i successful reconstructions for the i–th target. Sampling the identical sequence (or sequences which yield identical PAF_A) multiple times is clearly detrimental to our goal, and policy diversity is crucial to efficiently determine whether a given target autocorrelation is attainable. For this reason, we regularize the objective with a KL term between the policy and a uniform prior over legal completions of a partial trajectory, with the trained policy defined as the conformal shift of π_0:

$$\pi_0(a|s) := \binom{R - a - 2}{L - 2} \Big/ \binom{R - 1}{L - 1} \,, \quad \pi_\varphi(a|s) = \pi_0(a|s) \exp(-f_\varphi(s; a)) \,.$$

Here the policy head f_φ is a standard autoregressive transformer module. Written over the entire sampled trajectory, the objective we are implicitly maximizing with PPO reads as,

$$J(\theta) = \mathbf{E}_\pi \left[\sum_t r_t\right] - \alpha \cdot \mathsf{KL}\left(P_\pi \| P_0\right) = \mathbf{E}_\pi \left[\sum_t \left(r_t - \alpha \left(\log \pi_\varphi(a_t|s_t) - \log \pi_0(a_t|s_t)\right)\right)\right] \,. \tag{31}$$

We found this simple regularization to be significantly more sample–efficient than standard PPO, which tended to have a high pass@1 on training targets but very low pass@k–statistics on the validation set — the actor would place an inordinate amount of probability mass on a relatively small number of feasible sequences A with many possible completions. We describe the architecture used for this baseline model in Appendix A.

Table 1: Increase in sampling efficiency due to the trained policy. Note the advantage grows with the order.

n	Random policy	Arithmetic-free / rand. (↑)	Arithmetic / rand. (↑)
103	$(1.35 \pm 0.04) \times 10^{-7}$	1,025×	**7,407**×
111	$(1.82 \pm 0.08) \times 10^{-8}$	1,691×	**13,671**×
115	$(6.2 \pm 0.6) \times 10^{-9}$	2,309×	**22,888**×

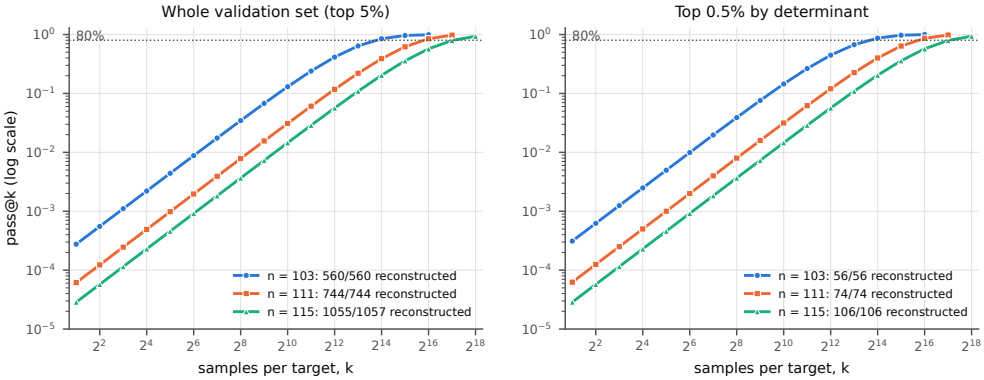

Figure 5: pass@k curves for the final trained policy, computed using unbiased estimator for ρ_k (30).

3.5 Results

For orders $n = 103, 111, 115$, the policy is able to realize the sequences (A, B) corresponding to the highest determinant predicted, $\max_p \det X_0 \prod_{t>0} \hat{p}_t$, via an exhaustive scan of the autocorrelation targets $p(t)$. The results for the record–setting sequences and the efficiency of our scan on the validation set (the highest–$|\det X|$ 5% of targets) are reported below in Table 2. Note that not every target autocorrelation may be guaranteed to admit a realizing pair (A, B), and, while being close to 1, the coverage reported on the validation set should be interpreted as a lower bound on the realizable fraction.

From the pass@k rate evaluated for the entire validation set and the top–end, exhibited in Table 4, we observe the difficulty of reconstruction is relatively insensitive to the determinant of the predicted target. Further, it appears each increment of $+4$ in the order roughly corresponds to a two–fold increase in reconstruction difficulty.

To demonstrate the utility of RL in this setting, we randomly sample permissible sequences A (with k_A positive elements) and run a Monte Carlo estimate of the pass@1 rate for this 'random' policy, with a fixed number of samples per target at each order 2^{24} at $n = 103$ and 2^{25} at $n = 111, 115$, and exhibit the results in Table 1. The trained policy increases the sampling efficiency by over three orders of magnitude for each n.

4 Conclusions

In this work, we have developed a framework for Hadamard maximal determinant problem which reformulates the determinant maximization problem as reinforcement-learning-based PAF-sum target reconstruction. The resulting constructions allow us to produce matrices in order $n = 51$ and orders from $n = 103$ to $n = 119$ for $n \equiv 3 \pmod 4$ exceeding previously known largest determinants.

Although the size of state space ranges from 10^{13} to 10^{33}, decomposing reconstruction into arithmetic feasibility and compatibility across different projections into cyclic groups enables reinforcement-learning to develop search strategies that generalize to previously unseen PAF-sum targets. Furthermore, while classical methods, such as simulated annealing, restart the search for each new target, the learned policy retains experience, potentially reducing reconstruction time.

Ethics Statement

This research adheres to ICLR Code of Ethics. This work concerns mathematical and computational methods for finding maximal determinant matrices. It does not involve human subjects, or personal data. We do not anticipate any specific societal harms arising from the methods or results presented here.

Reproducibility Statement

We will release all relevant code and supporting data upon acceptance.

The use of Large Language Models

We used large language models to help structure and polish writing, to identify relevant literature, discuss methodological choices and to optimize code implementing some of the search methods. The authors verified LLM outputs and take responsibility for the final manuscript, mathematical claims, results, and code.

References

Forest Agostinelli, Stephen McAleer, Alexander Shmakov, and Pierre Baldi. Solving the rubik's cube with deep reinforcement learning and search. *Nature Machine Intelligence*, 1(8):356–363, 2019.

J. A. Armario and D. L. Flannery. Almost supplementary difference sets and quaternary sequences with optimal autocorrelation. *Cryptography and Communications*, 12(4):757–768, 2020. doi: 10.1007/s12095-019-00418-x.

Joel Brenner. The hadamard maximum determinant problem. *The American Mathematical Monthly*, 79(6):626–630, 1972. ISSN 00029890, 19300972. URL http://www.jstor.org/stable/2317092.

Richard P. Brent and Adam B. Yedidia. Computation of maximal determinants of binary circulant matrices. *Journal of Integer Sequences*, 21(5):Article 18.5.6, 2018. ISSN 1530-7638. Scopus EID: 2-s2.0-85048808645.

Richard P Brent, William Orrick, Judy-anne Osborn, and Paul Zimmermann. Maximal determinants and saturated d-optimal designs of orders 19 and 37. *arXiv preprint arXiv:1112.4160*, 2011.

Patrick Browne, Ronan Egan, Fintan Hegarty, and Padraig Ó Catháin. A survey of the hadamard maximal determinant problem. *The Electronic Journal of Combinatorics*, 28(4):P4.41, December 2021. doi: 10.37236/10367. URL https://www.combinatorics.org/ojs/index.php/eljc/article/view/v28i4p41.

Yuri Burda, Harrison Edwards, Amos Storkey, and Oleg Klimov. Exploration by random network distillation. In *International Conference on Learning Representations (ICLR)*, 2019.

Felix Chalumeau, Shikha Surana, Clément Bonnet, Nathan Grinsztajn, Arnu Pretorius, Alexandre Laterre, and Thomas D. Barrett. Combinatorial optimization with policy adaptation using latent space search. In *Advances in Neural Information Processing Systems*, volume 36, pp. 7947–7959, 2023. URL https://arxiv.org/abs/2311.13569.

Dragomir Z Djokovic and Ilias S Kotsireas. New results on d-optimal matrices. *arXiv preprint arXiv:1103.3626*, 2011.

Hartmut Ehlich. Determinatabschätzungen für binäre matrizen. *Mathematische Zeitschrift*, 83:123–132, 1964a. URL http://eudml.org/doc/170219.

Hartmut Ehlich. Determinatenabschätzung für binäre matrizen mit $n \sim 3 \bmod 4$. *Mathematische Zeitschrift*, 84:438–447, 1964b. URL http://eudml.org/doc/170280.

Jacques Hadamard. Rèsolution d'une question relative aux dèterminants. *Bulletin des Sciences Mathématiques*, 2(17):240–246, 1893.

Hyeonah Kim, Minsu Kim, Sungsoo Ahn, and Jinkyoo Park. Symmetric replay training: Enhancing sample efficiency in deep reinforcement learning for combinatorial optimization. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 24110–24136. PMLR, 2024. URL https://proceedings.mlr.press/v235/kim24o.html.

Wouter Kool, Herke van Hoof, and Max Welling. Attention, learn to solve routing problems! In *International Conference on Learning Representations*, 2019. URL https://arxiv.org/abs/1803.08475.

Ilias S. Kotsireas, Youtong Liu, and Jing Yang. Paf reconstruction with the orbits method. In *Computer Algebra in Scientific Computing*, volume 11661 of *Lecture Notes in Computer Science*, pp. 315–329. Springer, Cham, 2019. doi: 10.1007/978-3-030-26831-2_21.

Ilias S Kotsireas, Christoph Koutschan, and Arne Winterhof. Quaternary legendre pairs ii. *Discrete Mathematics*, 348(9):114501, 2025.

Yeong-Dae Kwon, Jinho Choo, Byoungjip Kim, Iljoo Yoon, Youngjune Gwon, and Seungjai Min. POMO: Policy optimization with multiple optima for reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 33, pp. 21188–21198, 2020. URL https://arxiv.org/abs/2010.16011.

William Orrick. The Hadamard maximal determinant problem, October 2012. URL http://www.indiana.edu/~maxdet/. Archived web resource.

William P Orrick. The maximal {-1, 1}-determinant of order 15. *Metrika*, 62(2):195–219, 2005.

Ram Ramrakhya, Dhruv Batra, Erik Wijmans, and Abhishek Das. Pirlnav: Pretraining with imitation and rl finetuning for objectnav. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17896–17906. IEEE, 2023.

Jean-Pierre Serre. *Local Fields*, volume 67 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, Heidelberg, Berlin, 1979. ISBN 978-1-4757-5675-3. doi: 10.1007/978-1-4757-5673-9.

Jean-Pierre Serre. *A course in arithmetic*. Springer, 1993.

The PARI Group. *PARI/GP, Version 2.17.2*. Université de Bordeaux, Bordeaux, France, 2025. https://pari.math.u-bordeaux.fr/.

The SageMath Community. CyPari2: A python interface to PARI/GP, version 2.2.4. Computer software, 2025. https://github.com/sagemath/cypari2.

Geordie Williamson, Oded Yacobi, and Paul Zinn-Justin. Generating hadamard matrices with transformers. *arXiv preprint arXiv:2604.11101*, 2026.

A Arithmetic–free baseline architecture

Our architecture for the arithmetic–free baseline is a small autoregressive transformer decoder D_φ, which, in addition to an embedding of the previous action a_{t-1}, accepts as input the Fourier transform $\hat{p}_k$ of the target sequence, as well as an embedding of the last position committed c_{t-1} and the remaining gaps to emit, $k_A - 1 - t$. The decoder consists of 3 attention blocks, $d = 192$, and 6 heads of width 32. D_φ outputs logits corresponding to a sequence of valid gaps between consecutive $+1$ elements of A; $D_\varphi : \mathbb{R}^{(h+2)\times d} \to \mathbb{R}^T$, where $T = k_A - 1$ is the trajectory length. The policy is an MLP head over the output of the decoder. The dataset for each order consists of all possible enumerated periodic autocorrelation targets $p(t)$ for that order, of size $O(10^5)$. We order the dataset by determinant and train on the bottom 95% of sequences, retaining the top 5% of sequences for evaluation. The policy is trained using PPO without an entropy term in the objective, with $\alpha = 0.03$ set after a sweep at $n = 103$. The critic is an MLP head over the objective and shares the same decoder trunk as the policy head. The architecture is held identical across all orders apart from the necessary changes to input dimensions and embedding tables.

Table 2: The record-achieving matrices at orders $n = 2m + 1 = 103, 111, 115$. $\sigma.$ denotes the row sum, and $|\det X|/D_{\text{ref}}$ is the exact determinant over the best previously known determinant at that order Orrick (2012).

$$X(A,B) = \begin{pmatrix} 1 & j^T & -j^T \\ j & A & B \\ -j & B^T & -A^T \end{pmatrix}, \qquad A = \mathsf{Circ}(a), \quad B = \mathsf{Circ}(b), \qquad \mathsf{Circ}(x)_{ij} = x_{(j-i) \bmod m}$$

	First row	$\sigma.$
$n = 103,\ m = 51,\ \lvert\det X\rvert/D_{\text{ref}} = 1.0000004$		
a_{103}	−−++++++++− −+−−−+−++− −−++−+−+−− +−+−++−−+− ++−−+−+++− −	1
b_{103}	++−−+−−+−+ −−+++−+−+− −−+−++−−++ −−++−−−−++ +++++++++− +	7
$n = 111,\ m = 55,\ \lvert\det X\rvert/D_{\text{ref}} = 1.0126294$		
a_{111}	++++−−++−+ −−−+−++−+− +++++−−++− −+−−−++−−+ −+−−−++++− −+−−−	1
b_{111}	+++−+−−+−− ++−−+++−+− −−−++−+−+− +−++−+−−−− −++++−++++ −++++	7
$n = 115,\ m = 57,\ \lvert\det X\rvert/D_{\text{ref}} = 1.0182876$		
a_{115}	++++−−−−+− −++−+−+−−+ −+−−+++−−− +−+++−+−−+ −−++−−+−++ +−+++−−	1
b_{115}	++−−−−+−+− −++++−+−+− −−++−+++−+ −−−−++−++− −−+++−+−++ +++++++	7

Table 3: Record determinant sample statistics. 'Distinct pairs' indicates pairs which are not related by a trivial joint symmetry to any other pair. 'Samples needed' indicates the cumulative number of samples needed up to and including the first sampled batch which yielded a reconstruction. The wallclock time is measured as the target's share of a batched run, i.e. (number of samples needed) $\times$ (average seconds per sample).

n	$\det X/$ record	Recon-structions	Distinct pairs	Samples needed	Samples drawn	Time to first (s)	Total time (s)
103	1.0000004	61	8	$\leq$ 17,408	148,480	1.2	10.4
111	1.0126294	7	3	$\leq$ 74,752	271,360	18.9	66.7
115	1.0182876	1	1	$\leq$ 403,456	534,528	44.0	56.9

B Further results

B.1 New record matrices

Here we exhibit new record–determinant matrices obtained by reinforcement learning in orders $n = 103, 111, 115$. All matrices are of the bordered bicirculant form defined by the *uncorrelated* construction (1), so we elide the border, and only report the circulant component of the matrices $A_n = \text{circ}(a_n)$ and $B_n = \text{circ}(b_n)$ for each order n.

B.2 Dataset construction

To generate the dataset used for the arithmetic–free method, note the orders are sufficiently small to exhaustively enumerate all possible autocorrelation functions[1]. We restrict ourselves to targets which yield a determinant of 80% or higher for each order, which yields an $O(10^4)$ dataset $\mathcal{D}_n = \{p_n(t)\}$ of autocorrelation functions at all orders $n \equiv 3 \pmod 4$ we consider, which form a superset of the feasible autocorrelation functions admitting the decomposition (5). Note we have recast the problem into a form that is nominally insensitive to the determinant of the bicirculant matrix X constructed from the pair (A, B). Here we exhibit further results showcasing the efficiency of a trained policy over a random baseline.

[1] $p(t)$ is a sequence of length $m = (n-1)/2$ with reflection symmetry along the midpoint, so one only needs to enumerate all $(m-1)/2$–sequences of binary strings, compute their orbit under the permutation $p \mapsto p \circ \lambda$ for some multiplicative index shift $\lambda \in \mathbb{N}$, and record the canonical representative of this orbit as the lexicographic minimum.

Table 4: Pair (A, B) reconstruction statistics on the validation set (the top 5% of targets measured by determinant) and on its top end (the top 0.5%). 'Fraction reconstructed' counts targets with at least one certified pair over all samples drawn. 'Samples/target' is the largest number drawn on any one target. pass@k is the mean over targets drawn from the respective sets of the unbiased estimator ρ_k.

n	Fraction reconstructed	Samples/target	pass@1	pass@256	pass@1024	pass@2^{12}	pass@2^{16}
Validation set (top 5%)							
103	100% (560/560)	$\leq 279{,}552$	1.37×10^{-4}	0.0336	0.126	0.405	0.992
111	100% (744/744)	$\leq 402{,}432$	3.07×10^{-5}	0.00783	0.0309	0.118	0.849
115	99.7% (1055/1057)	$\leq 2{,}500{,}608$	1.42×10^{-5}	0.00367	0.0146	0.0567	0.572
Top 0.5%							
103	100% (56/56)		1.81×10^{-4}	0.0388	0.145	0.459	0.997
111	100% (74/74)	(same as above)	3.16×10^{-5}	0.0083	0.0328	0.125	0.898
115	100% (106/106)		1.35×10^{-5}	0.00371	0.0147	0.0572	0.577

C Further results on PAFs

Proposition C.1. *Let N be the number of times $p(s) = +2$ for $s \neq 0$, then:*

$$|a|^2 + |b|^2 = 4N + 2\,. \tag{32}$$

Proof. The proof of the original statement can be found in Djokovic & Kotsireas (2011). Let N be the number of times $p(s) = 2$ for $s \neq 0$, then:

$$\sum_{s=0}^{m-1} p(s) = 2m + 2N - 2(m - 1 - N) = 4N + 2\,. \tag{33}$$

Using the definition of the periodic autocorrelation functions (4), we see that:

$$\sum_{s=0}^{m-1} \mathrm{PAF}_a(s) = \sum_{s=0}^{m-1} a_s^2 + \sum_{j=0}^{m-1}\sum_{s=1}^{m-1} a_j a_{j+s} = \sum_{s=0}^{m-1} a_s^2 + 2\sum_{j<k} a_j a_k = \left(\sum_{j=0}^{m-1} a_j\right)^2 \tag{34}$$

Let $|a|$, $|b|$ denote the row sum of a, b, respectively. Combining (33) and (34) yields $|a|^2 + |b|^2 = 4N + 2$. □

C.1 Correlated $X^{\mathrm{corr.}}$

We follow the same exact procedure for $X = X^{\mathrm{corr.}}(A, B)$. Except, the Gram matrix G of X is given by:

$$G = XX^T = \begin{pmatrix} n & \gamma^T & \gamma^T \\ \gamma & C & j \otimes j - K \\ \gamma & j \otimes j - K^T & C \end{pmatrix}, \tag{35}$$

where $\gamma = (1+|a|-|b|)j$ and C is same as before. The main structural difference is the off-diagonal contribution coming from $K = A(B+B^T)$. Following Ehlich Gram constraint, we impose $K_{ij} = 2$ condition on A and B. While it is impossible to rewrite this condition purely in terms of PAF-sum targets $p(s)$, we may still write it in terms of the constraints on the cross-correlations of a and b. In particular, by taking the Fourier transform of $K = 2j \otimes j$ it is easy to show that the relevant constraints are:

$$\text{Zero mode: } |a||b| = m, \quad \text{Higher modes: } \widehat{a}_k \mathrm{Re}(\widehat{b}_k) = 0\,, \tag{36}$$

where latter may be rewritten purely in terms of constraints on the convolution of sequences: $a * c = 2|a||b|/m = 2$, where $c_t := b_t + b_{-t \pmod m}$. Assuming such sequences exist, we calculate the

determinant in the same manner as before: the Fourier modes of X are given by:

$$\widehat{X}_0 = \begin{pmatrix} 1 & \sqrt{m} & -\sqrt{m} \\ \sqrt{m} & |a| & |b| \\ \sqrt{m} & -|b| & -|a| \end{pmatrix}, \quad \widehat{X}_{t>0} = \begin{pmatrix} \widehat{a}_t & \widehat{b}_t \\ -\widehat{b}_t & -\widehat{\bar{a}}_t \end{pmatrix}. \tag{37}$$

While $\det \widehat{X}_0$ similarly only depends on the segment sums $|a|$ and $|b|$, the form of the determinant of the higher modes $\det \widehat{X}_{t>0}$ is given by:

$$|\det \widehat{X}_{t>0}|^2 = \left|\widehat{b}_t^2 - |\widehat{a}_t|^2\right|^2 = (|\widehat{a}_t|^2 + |\widehat{b}_t|^2)^2 - 4(|\widehat{a}_t|\mathrm{Re}(\widehat{b}_t))^2 = (|\widehat{a}_t|^2 + |\widehat{b}_t|^2)^2 \tag{38}$$

where the third equality holds true due to assumption $K = 2j \otimes j$. Using the same expansions of higher mode determinants as before, we obtain the same exact expression for the determinant of the higher modes of X, while zero mode can similarly be determined from the segment sums. The exhaustive list of all admissible targets for $n = 51$ parametrized using $X^{\mathrm{corr.}}$ is shown in Fig. 2.

D Numerical and RL approaches

Order n	Segment sums $(|a|, |b|)$	$\log_{10} |\mathcal{S}(|a|, |b|, (n-1)/2)|$
51	$(5, 5)$	13.03
103	$(1, 7)$	28.59
107	$(1, 7)$	29.78
111	$(1, 7)$	30.98
115	$(1, 7)$	32.17
119	$(1, 7)$	33.36

Table 5: Size of the state space $\mathcal{S}(|a|, |b|, (n-1)/2)$ for orders n where we search for $n = 51$ and $103 \leq n \leq 119$ with $n \equiv 3 \pmod 4$.

D.1 Order 51 using $X^{\mathrm{corr.}}(A, B)$

Simulated annealing with energy functional E given by:

$$E(a, b) := H(a, b) + \sum_{t=0}^{m-1} \left|(a * c)_t - 2\right|^2 \tag{39}$$

Where $c_t = b_t + b_{-t \pmod m}$ and $H(a, b)$ is (15). We reconstructed all 10 record breaking PAF-sum targets at order 51. The reconstructions are summarized in the Table 6.

Sequence a	Sequence b	PAF-sum target
`--+-+-+---++++++-++-++-++`	`+++++++-+++--+++--+-+----`	`+2 +2 +2 +2 +2 +2 -2 -2 -2 -2 -2 -2`
`-+++-+++-++-+-----++++-++`	`+----++-+-+++--++-+-+++++`	`+2 +2 +2 -2 -2 -2 -2 -2 -2 +2 +2 +2`
`++-+-+++-----+-++++++--+++`	`++++-++--+++-+-+-++-++---`	`+2 +2 -2 -2 -2 -2 +2 +2 +2 +2 -2 -2`
`++++-++++--+--++-++++---+`	`++-+-+---++--+++-++++-+-`	`+2 -2 -2 +2 +2 -2 -2 +2 +2 -2 -2 +2`
`+++----++++++-++--++-+-+-`	`+++--+-+--+--+++++-++++--`	`+2 -2 -2 -2 +2 +2 +2 -2 -2 -2 +2 +2`
`+-++-+++-+-+----++++-++`	`+++-+++--++++--+-++-+-+--`	`-2 +2 -2 +2 +2 -2 +2 -2 +2 -2 +2 -2`
`+++-+-+++--++-++--+-+--++`	`+----+-+-++--+++-+-++++++`	`-2 +2 -2 +2 -2 +2 -2 +2 -2 +2 -2 +2`
`+++++-++--+++-+---+-+--++`	`++-+-+++--+--++++--++-+-`	`-2 -2 +2 +2 -2 -2 +2 +2 -2 +2 +2 -2`
`+----++-++--+-+-+++-+++++`	`+---++--+-++-+-++-+++-+++`	`-2 -2 +2 -2 +2 +2 -2 +2 +2 -2 +2 -2`
`+-++++--++++-+-+++--+---`	`++---+-++-++-+-++--++++-`	`-2 -2 +2 -2 -2 +2 +2 -2 +2 +2 -2 +2`

Table 6: Exact reconstructions for all 10 maximal determinant PAF-sum targets for order $n = 51$.

We limited experimentation to simulated annealing in this order as it was able to produce solutions almost instantaneously.

D.2 Orders 107 and 119 using simulated annealing

Since $m = (n-1)/2$ for both orders 107 and 119 is prime, the lifting setup described in Sec 3.2 does not apply. However, we still perform simulated annealing on energy functional $E(a, b) := H(a, b)$ (15). We report reconstruction of a maximal determinant PAF-sum target for each of the orders 107 and 119 in Table 7.

Order n		Sequences	PAF-sum target
107	a	`-+-+-++++-+--++-+-++++--++++-++--------+-++-++-----++`	`-2 +2 -2 +2 -2 -2 -2 -2 -2 -2 -2 -2 -2`
	b	`++-+-+---++++-+-++++-++-+++---+---++--++-++-+++-+---+`	`-2 +2 -2 +2 -2 +2 -2 +2 -2 -2 -2 -2 -2`
119	a	`-----+---+-++-+-+-+--+--++++++-+-++-++----++-+--++--++-+++`	`-2 -2 -2 +2 -2 -2 -2 +2 -2 -2 -2 -2 +2 -2 -2`
	b	`++++++++---+-+++--+++--++-+-++++---+--++-+-+----++-+---+++-`	`-2 +2 -2 -2 -2 +2 -2 -2 -2 +2 -2 -2 -2 -2`

Table 7: Exact reconstructions of a maximal determinant PAF-sum target for each order $n \in \{107, 119\}$.

D.3 Order 103 using RL on full state space $\mathcal{S}(|a|, |b|, m)$

From the exhaustive enumeration (See Fig. 2), PAF-sum targets in $n \equiv 3 \pmod 4$ dimensions from 103 to 119 whose determinant exceeds the previous record Orrick (2012) come from $X(A, B)$ family. For all of these matrices the segment sums attaining the maximal determinant are $(1, 7)$. Since these matrices come from $X(A, B)$, in order for a pair of sequences (a, b) to realize those targets, they must solve only one constraint:

$$H(a, b) = 0 \tag{40}$$

where $H(a, b)$ is the PAF-sum objective defined in (15). In this section, we focus specifically on order 103 and describe the training algorithm of an agent which takes an arbitrary PAF-sum target as an input and produces its reconstruction. Motivated by the results of Ramrakhya et al. (2023); Agostinelli et al. (2019), we first sample a small-scale PAF-sum reconstruction dataset (Alg. 1), pretrain the agent on it using behavior cloning (Alg. 2), and then finetune it through a curriculum of increasing scramble depth (Alg. 3) to allow it to generalize to reconstruct unseen PAF-sum targets. Conceptually, the illustration of this transfer and generalization to interesting PAF-sum targets is shown in Fig. 1. We pick $\mathcal{D}_{\text{val}}$ to consist of top 20% of all enumerated PAF-sum targets in Fig. 2. At inference, the trained agent attempts to reconstruct these targets from uniformly sampled initial states (Alg. 4). The PAF-sum targets $\mathcal{D}_{\text{PAF}}$ used for pretraining and finetuning do not overlap with $\mathcal{D}_{\text{val}}$, moreover, it is highly unlikely that it will contain reconstruction of the maximal PAF-sum target. This is mainly because constructing an admissible PAF-sum target is a significantly easier problem than reconstructing a specific PAF-sum target. Figure 6 compares reconstruction rates for energy-guided, imitation learning and GRPO fine tuning for a k random swap scramble. We observe that at scramble depth 10-12 the imitation learning policy is unable to reconstruct the PAF-sum making it unfeasible for constructing full solutions.

D.4 Pretraining and Curriculum algorithms

Algorithm 1 Sampling pretraining PAF data

Require: Order $n \equiv 3 \pmod 4$, $m = (n-1)/2$. Pretraining dataset size $N > 0$.

I: Sampling pretraining data

$\mathcal{D}_{\text{PAF}} \leftarrow \emptyset$
while $|\mathcal{D}_{\text{PAF}}| < N$ **do**
  $(a, b) \leftarrow \textsc{SimulatedAnnealing}(m)$
  $p \leftarrow \text{PAF}_a + \text{PAF}_b$
  if $\forall s \neq 0 \colon p(s) \in \{-2, +2\}$ **then**
    $\mathcal{D}_{\text{PAF}} \leftarrow \mathcal{D}_{\text{PAF}} \cup \{(a, b)\}$
  end if
end while

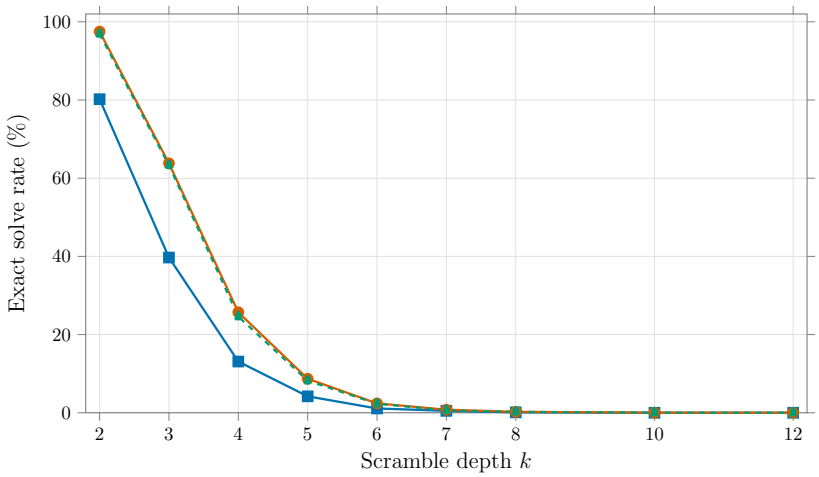

Figure 6: Exact reconstruction rate of non-arithmetic energy-based RL on the full state space $\mathcal{S}(|a|, |b|, m)$ for held-out order-103 targets after k random steps, with a budget of $2k$ moves. Energy-guided search (blue), imitation learning (orange), and GRPO fine-tuning (green, dashed), all decline sharply as scramble depth increases

Algorithm 2 Pretraining the agent

Require: Order $n \equiv 3 \pmod 4$. Pretraining dataset $\mathcal{D}_{\text{PAF}}$. Max scrambles $K_{\max} > 0$. Swap-action environment env.

$\mathcal{D}_{\text{traj}} \leftarrow \emptyset$
for $(a, b, p) \in \mathcal{D}_{\text{PAF}}$ **do**
  $s \leftarrow (a, b)$
  for $k = 1, \ldots, K_{\max}$ **do**
    $(s_0 = s, s_1, \ldots, s_k) \leftarrow \textsc{RandomScramble}(s, k, \text{env})$
    $\tau \leftarrow \textsc{ShortestPath}(s_0, s_k)$
    $\mathcal{D}_{\text{traj}} \leftarrow \mathcal{D}_{\text{traj}} \cup \{(\tau, p)\}$
  end for
end for
$\theta \leftarrow \text{BC}(\pi_\theta, \mathcal{D}_{\text{traj}})$ ▷ Behavior Cloning

Algorithm 3 Curriculum learning over increasing scramble depth

Require: Order $n \equiv 3 \pmod 4$. Pretraining dataset $\mathcal{D}_{\text{PAF}}$. Swap-action environment env. Curriculum scheduler k_t.

for $t = 1, \ldots, T_{\max}$ **do**
  $(p, s) \leftarrow \mathcal{D}_{\text{PAF}}$
  $\epsilon \sim \mathcal{N}(0, 1)$
  if $\epsilon > \epsilon_{\text{th}}$ **then**
    $(\ldots, s_k) \leftarrow \textsc{RandomScramble}(s, k_t, \text{env})$
  else
    $s_k \sim \mathcal{U}(\text{env})$ ▷ Random uniform from state space.
  end if
  $\tau \leftarrow \textsc{Rollout}(\pi_\theta, s_k, p, \text{env})$
  $\theta \leftarrow \textsc{RLOptimize}(\theta, \tau, p)$
end for

Algorithm 4 Inference / generation

Require: Order $n \equiv 3 \pmod 4$. Max scrambles $K_{\max} > 0$. Swap-action environment env. Trained policy π_θ.

Inference / generation
$S \leftarrow \emptyset$
for $p \in \mathcal{D}_{\text{val}}$ **do**
  $s_0 \leftarrow \mathcal{U}(\text{env})$
  $s^\star \leftarrow \textsc{Rollout}(\pi_\theta, s_0, p, \text{env})$
  if $H(s^\star) = 0$ **then**
    $S \leftarrow S \cup \{(s^\star, p)\}$
  end if
end for
return π_θ, S

E ARITHMETIC OBSTRUCTIONS

The key identification is that we may regard sequences $\{a_j\}_{j=0}^{m-1}$ as elements of $\mathbb{Q}[C_m]$ where C_m is a cyclic group of order m. This therefore allows us to interpret PAF functions of a as products:

$$aa^* = \sum_{j=0}^{m-1} \mathrm{PAF}_a(j) z^j \in \mathbb{Q}[z]/(z^m - 1) \cong \mathbb{Q}[C_m] \tag{41}$$

Therefore, the problem of PAF reconstruction of some profile r is essentially a problem of searching for $a \in \mathbb{Q}[C_m]$ such that $aa^* = r$, where r here is implicitly identified as an element of $\mathbb{Q}[C_m]$. While this reformulation doesn't necessarily simplify the problem, we may look for local obstructions by projecting a to the cyclotomic subfields $\mathbb{Q}(\zeta_d)$ of the decomposition:

$$\mathbb{Q}[C_m] = \prod_{d|m} \mathbb{Q}(\zeta_d)\,, \tag{42}$$

Theorem E.1. *Let $d \mid m$ for some $d > 2$ and K_d^+ be the maximal real subfield of $K_d := \mathbb{Q}(\zeta_d)$, then $b_d b_d^* = r_d \in (K_d^+)^\times$ has a solution $b_d \in K_d$ iff:*

$$[r_d] = 0 \text{ in } H^2(K_d/K_d^+)\,, \tag{43}$$

where $H^2(K_d/K_d^+)$ is the Galois cohomology $H^2(\mathrm{Gal}(K_d/K_d^+), K_d^\times)$.

Proof. Since $H^2(K_d/K_d^+) \simeq (K_d^+)^\times / N_{K_d/K_d^+}(K_d^\times)$ Serre (1979). Let c be the generator of $\mathrm{Gal}(K_d/K_d^+) \simeq \mathbb{Z}/(2)$, then we have:

$$[r_d] = 0 \iff r_d = N_{K_d/K_d^+}(b) \tag{44}$$

for some $b_d \in K_d$, thus $N_{K_d/K_d^+}(b_d) = c(b_d) b_d = b_d b_d^* = r_d$. □

The case $r_d = 0$ can be handled separately, as it has unique solution $b_d = 0$.

We may infer existence of global obstruction $[r_d] \neq 0$ by doing finitely many local tests. In particular, we have:

Theorem E.2. *Let $\Delta_d := (\zeta_d - \zeta_d^{-1})^2$. Let r_d be such that it passes the Archimedean test: $\sigma(r_d) > 0$ for all embeddings $\sigma\colon K_d^+ \to \mathbb{R}$. Then, we have the following equivalence:*

$$\forall 0 \neq v \in \mathrm{Spec}\, \mathcal{O}_{K_d^+}\colon (r_d, \Delta_d)_v = 1 \iff [r_d] = 0 \text{ in } H^2(K_d/K_d^+)\,, \tag{45}$$

where $(r_d, \Delta_d)_v$ is a Hilbert symbol at a place v.

Proof. The element $\zeta_d \in K_d$ has a minimal polynomial over K_d^+ given by $f_d = X^2 - s_d X + 1 = 0$, where $s_d = \zeta_d + \zeta_d^{-1}$. The discriminant of f_d is $\Delta_d = s_d^2 - 4 = (\zeta_d - \zeta_d^{-1})^2$, thus $K_d = K_d^+(\sqrt{\Delta_d})$. From Serre (1993), the condition $(r_d, \Delta_d)_v = 1$ is equivalent to $r_d \in N_{(K_d)_v^+(\sqrt{\Delta})/(K_d)_v^+}((K_d)_v^+(\sqrt{\Delta_d})^\times)$, i.e. r_d is a local norm. Since $(r_d, \Delta_d)_v = 1$ at all finite places v and $\sigma(r_d) > 0$ for all $\sigma\colon K_d^+ \to \mathbb{R}$, by Hasse norm theorem, r_d must also be global norm. Converse is obvious. □

We use cypari2 The SageMath Community (2025); The PARI Group (2025) to efficiently calculate Hilbert symbols (45) at all relevant places $v \in \mathrm{Spec}\, \mathcal{O}_{K_d^+}$.