

# UNIVERSAL RANK TESTS FOR NONCOMMUTATIVE RATIONAL FORMULAS

MICHAEL SCHROEDER  
INDEPENDENT RESEARCHER

**ABSTRACT.** Testing noncommutative rational expressions requires preserving the domains of their inverse operations while detecting their algebraic rank. We construct universal matrix evaluation lists that achieve both over every field, extending polynomial rational testing constructions over  $\mathbb{Q}$  to arbitrary coefficient fields.

For  $n$  variables, total tree size  $s$ , and inversion height  $h$ , the list depends only on these bounds and the characteristic. Every family of individually admissible rational expression matrices within these bounds has a common evaluation in the list at which every original inverse is defined and every matrix attains its maximum normalized rank. All evaluation entries lie in the prime subfield.

In characteristic two, the matrix dimension is  $O(s^2 \log^3(s + 2))$ , and the list length is  $O((h + 1)s^8 \log^6(s + 2) \log(n + 2))$ . In every other characteristic, the corresponding bounds are  $O(s^4 \log^4(s + 2))$  and  $O((h + 1)s^{10} \log^7(s + 2) \log(n + 2))$ . Thus matrix dimension is independent of the number of variables and inversion height, while list length depends only logarithmically on the number of variables. The construction yields coefficient-independent, nonadaptive tests for affine noncommutative rank over every field.

One deterministic Turing machine constructs the complete densely encoded lists in polynomial time, with the characteristic supplied in binary. We prove the main theorem, including serialization and the machine's runtime bound, in Lean.

## 1. INTRODUCTION

Matrix evaluations turn questions about noncommutative expressions into finite linear-algebra computations. Variables are replaced by square matrices, coefficients by scalar matrices, and the expression is evaluated operation by operation. Rational expressions introduce an additional requirement: every matrix to be inverted must be invertible. This condition belongs to the expression as written. Even  $0 \cdot x^{-1}$  requires an invertible value of  $x$ , although

---

*Date:* 18 October 2026.

*Key words and phrases.* Noncommutative rational formulas, polynomial identity testing, noncommutative rank, finite fields, derandomization, Lean formalization.

ORCID: [0009-0004-3249-0195](https://orcid.org/0009-0004-3249-0195).

DOI (reserved): [10.5281/zenodo.23453468](https://doi.org/10.5281/zenodo.23453468).

its value is then zero. A useful test must therefore preserve the domain as well as reveal the value or rank.

We seek a finite list chosen before the targets and their coefficients are known. It should work over every field of the chosen characteristic and provide one common evaluation preserving all domains and attaining all maximum ranks in a family. We measure both the matrix dimension and the number of evaluations.

Main result. Theorem 2.1 constructs such a list from the characteristic, number of variables  $n$ , total expression-tree size bound  $s$ , and inversion-height bound  $h$ . Its entries lie in the prime subfield. For every family within these bounds whose members are individually defined somewhere, one listed evaluation respects every original inverse and attains every member’s maximum normalized rank. Normalization divides the evaluated rank by the size of the substituted matrices; the maximum ranges over all dimensions and coefficient-field extensions. We prove that it is an integer and is attained.

The matrix dimension depends only on  $s$  and the characteristic. The list-length bound is polynomial in  $s$ , linear in  $h + 1$ , and logarithmic in  $n$ ; Table 1 gives the explicit bounds, including the sharper characteristic-two construction. One deterministic Turing machine produces the complete dense output in polynomial time, with the characteristic encoded in binary. The Lean proof covers this machine, its encoding and runtime, together with the domain and rank guarantees.

Corollaries 2.2 and 2.3 give scalar rational identity tests and universal tests for affine noncommutative rank. For the latter, the evaluation points are fixed before any rank-oracle answers. These tests are nonadaptive and independent of the coefficients. Proposition 2.4 shows why lists and a dependence on  $n$  are necessary in the stated regimes. The explicit dimensions remain large even for small inputs (Section 9); the gain is a polynomial complexity guarantee.

Relation to previous work. Classical results on rational identities and realizations provide the structural background [1, 16, 8]. Operator scaling and rank regularity yield white-box algorithms and small rank witnesses [14, 18, 17, 10]. For universal tests, Arvind et al. construct quasipolynomial rational hitting sets whose dimension is already independent of the number of variables [2]; Arvind and Joglekar give honest bivariate reductions [3].

The closest predecessor is OpenAI’s polynomial rational hitting-list construction over  $\mathbb{Q}$ , which also gives affine-rank tests [21, Corollary 6.1]. Its division-free counterparts give single hitting points in characteristic zero and positive characteristic [20, 23]. OpenAI repository entry 116 records the formalization of the September division-free and rational results [22]. Our contribution combines polynomial lists with strict-domain, simultaneous rank attainment over arbitrary coefficient fields, variable-independent dimension, and one verified construction across all characteristics.

Proof idea. We first express the domain and rank requirements as invertibility conditions on affine matrix pencils, whose entries are affine linear

TABLE 1. Universal scalar tests. The last two rows also attain matrix-family ranks over every field. Here  $\lambda_t = \lceil \log_2(t + 2) \rceil$ ; poly denotes a polynomial bound.

| Construction      | Formulas /<br>coefficients           | Dimension            | List length                              |
|-------------------|--------------------------------------|----------------------|------------------------------------------|
| Arvind et al. [2] | Rational / $\mathbb{Q}$              | $s^{O(\log s)}$      | $(ns)^{\text{polylog}(s)}$               |
| OpenAI [20]       | Division-free / all<br>char. 0       | $O(ns^2)$            | 1                                        |
| OpenAI [21]       | Rational / $\mathbb{Q}$              | $\text{poly}(n, s)$  | $\text{poly}(n, s)$                      |
| OpenAI [23]       | Division-free / all<br>char. $p > 0$ | $O(n^3 s^6)$         | 1                                        |
| This paper        | Rational / $p = 2$                   | $O(s^2 \lambda_s^3)$ | $O((h + 1)s^8 \lambda_s^6 \lambda_n)$    |
| This paper        | Rational / $p \neq 2$                | $O(s^4 \lambda_s^4)$ | $O((h + 1)s^{10} \lambda_s^7 \lambda_n)$ |

For inversion height  $\theta$ , Arvind et al. give dimension  $s^{O(\theta)}$  and list length  $(ns)^{O(\theta^5 \log s)}$ . The displayed row applies their domain-preserving reduction to  $\theta = O(\log s)$  [2, Fact 1 and Theorem 2, arXiv v4].

expressions. A division-algebra lift establishes integral maximum ranks and common witnesses; each original inverse retains its own certificate.

Perfect hashing [6] and short conjugate words reduce the targets to two variables while preserving domain and rank witnesses (Sections 3 and 4).

Next we construct fixed matrices that bring every feasible pencil close enough to full rank for the division lift to make it invertible. This quotient design develops the augmented rank-certificate matrix of [7, Theorem 5.3, p. 37] using folded determinants [15, 12]. The key estimate counts boundary roots with full multiplicity, exploiting the roots identified in [5, ECCC version, Remark 6.4, p. 31]. Section 5 proves the count and rank bound.

A one-parameter matrix curve then makes every certificate determinant a nonzero polynomial of bounded degree. Summing root bounds over the family ensures one sample satisfies all conditions. Descent puts the output in the prime subfield; explicit arithmetic, decoding and serialization give the bit bounds (Sections 6–8). The lift, realization, descent and normalization extend [21]; the machine model and encoding follow its **RationalHitting** formalization [22]. All rank-design arguments are proved here; adapted arguments and code are credited and rechecked. Section 9 and the Formal Supplement give the precise Lean correspondence and reproduction procedure.

## 2. THE THEOREM AND ITS SEMANTICS

A scalar expression is an actual tree with scalar and variable leaves, binary addition and ordered multiplication, and unary inverse. Size counts every vertex. Its inversion height is the maximum, over root-to-leaf paths, of the number of inverse gates on the path. Evaluation is strict: every original

inverse must exist, even in  $0 \cdot x^{-1}$  or a cancelled pair. A rational expression matrix has such trees as its entries; its domain requires every entry tree to be defined. Its row and column dimensions are positive integers. Total size counts all entries with repetition. Coefficients are central scalars from an arbitrary field  $F$ .

**Definition 2.1** (Strict admissibility and operational rank). An expression or expression matrix is *admissible* if it is defined at a finite square-matrix tuple over some extension of  $F$ . All extension fields and all positive matrix dimensions are allowed. For an admissible  $a \times b$  expression matrix  $M$ , define its *operational rank*

$$(2.1) \quad \tau(M) = \sup_{E/F, d \geq 1, X \in \text{Dom}_{d,E}(M)} \frac{\text{rank}_E M(X)}{d}.$$

The supremum is taken over the set of attainable nonnegative rational ratios, a bounded subset of  $[0, \min(a, b)]$ .

Theorem 3.4 proves that this number is an integer and is attained. We do not assume this by identifying a strict expression with a simplified free-field representative.

Example in characteristic two. Over  $\mathbb{F}_2$ , the height-one expression  $f(x, y) = (xy + yx)^{-1}$  has no scalar domain, even over extensions, since  $ab + ba = 0$  for scalars. Take

$$U = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad V = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad P = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Direct multiplication gives  $U^2 = V^2 = P^2 = I_2$ ,  $PUP^{-1} = V$ , and  $UV + VU = I_2$ . Thus  $f(U, V) = I_2$  and  $\tau(f) = 1$ , attained at a mutually conjugate unit pair. Even  $0 \cdot f$  retains the requirement that  $xy + yx$  be invertible. This pair is chosen for the example; the theorem constructs its list independently of  $f$ . Operational rank allows every positive dimension; permitted dimensions serve only the auxiliary lifts. The Formal Supplement gives a checked normalization, decoding and serialization example over  $\mathbb{F}_2$ .

**2.1. Main theorem: universal strict-domain and rank tests.** Write  $\mathcal{P} = \{0\} \cup \{p \in \mathbb{N} : p \text{ is prime}\}$ ,  $K_0 = \mathbb{Q}$ ,  $K_p = \mathbb{F}_p$  for  $p > 0$ , and  $\Lambda(t) = \lceil \log_2(t + 2) \rceil$ . If  $F$  has characteristic  $p$ , let  $X^F$  denote entrywise transport along the canonical embedding  $K_p \hookrightarrow F$ . Encodings are specified in Section 9.1.

**Theorem 2.1** (Universal polynomial matrix tests over every characteristic). *There exist a deterministic Turing machine  $G$ , a function  $D : \mathcal{P} \times \mathbb{N}_{\geq 1} \rightarrow \mathbb{N}_{\geq 1}$ , and absolute positive integers  $C, k$  with the following properties. Given  $p \in \mathcal{P}$  in binary and  $n, s \geq 1$ ,  $0 \leq h \leq s$  in unary,  $G$  outputs a nonempty finite list*

$$\mathcal{H}_{p,n,s,h} = (X^{(1)}, \dots, X^{(L)}), \quad X^{(j)} \in \text{Mat}_{D(p,s)}(K_p)^n.$$

The input promises that a nonzero tag  $p$  is prime. Neither a target formula nor a coefficient field is an input to  $G$ .

- (1) **Uniform bit complexity and dimensions.** The running time and the length of the complete dense binary output are each at most

$$(2.2) \quad C(n + s + \Lambda(p) + 1)^k.$$

With  $D = D(p, s)$ , the output also satisfies

$$(2.3) \quad \begin{array}{c|c|c} \text{Characteristic} & D & L \\ \hline p = 2 & Cs^2\Lambda(s)^3 & C(h+1)s^8\Lambda(s)^6\Lambda(n) \\ p \in \mathcal{P} \setminus \{2\} & Cs^4\Lambda(s)^4 & C(h+1)s^{10}\Lambda(s)^7\Lambda(n) \end{array}$$

where each table entry is an upper bound. In particular, the actual dimension is independent of  $n$ ,  $h$ , and all target coefficients.

- (2) **One valid point attaining all maximum ranks.** Let  $F$  be any field of characteristic  $p$ . Let  $M_1, \dots, M_m$ ,  $m \geq 1$ , be individually admissible rational expression matrices over  $F$  in  $n$  variables, with positive shapes  $a_i \times b_i$ , total entry-tree size at most  $s$ , and every entry's inversion height at most  $h$ . There exist a single  $j \in \{1, \dots, L\}$  and natural numbers  $\rho_i \leq \min(a_i, b_i)$  such that all the original entry trees of all  $M_i$  are defined at  $(X^{(j)})^F$  and

$$(2.4) \quad \text{rank}_F M_i((X^{(j)})^F) = D\rho_i \quad (1 \leq i \leq m).$$

Moreover, for every field extension  $E/F$ , every  $d \geq 1$ , and every  $Z \in \text{Dom}_{d,E}(M_i)$ ,

$$(2.5) \quad \text{rank}_E M_i(Z) \leq d\rho_i.$$

The rank in (2.4) is the ordinary rank of the  $a_i D$ -by- $b_i D$  evaluated block matrix. The numbers  $\rho_i$  are therefore exactly the operational ranks  $\tau(M_i)$ .

- (3) **Structured outputs.** For every listed tuple, all its  $n$  coordinates are invertible and mutually conjugate over  $K_p$ .

All constants and the single machine are independent of the characteristic, coefficient field, family, and coefficient values.

The binary characteristic uses the self-delimiting code of Section 9.1. No behavior is required outside the stated input promises, including malformed encodings or composite positive tags.

An exact evaluation oracle returns either “undefined” or the full matrix value at a requested tuple. Testing that value for zero or invertibility uses separate field arithmetic; it is not charged to  $G$ .

**Corollary 2.2** (Scalar rational identity testing). *For every admissible scalar rational formula  $f$  over any field  $F$  of characteristic  $p$ , of size at most  $s$  and height at most  $h$ , the same list contains a point in the original domain of  $f$ .*

Furthermore,

$f$  has a nonzero defined evaluation over some extension of  $F$   
 $\iff$  some listed point is defined for  $f$  and gives an invertible value.

Thus an evaluation oracle that distinguishes undefined values can test identity using this fixed list. Setting  $h = s$  permits every inverse nesting compatible with size  $s$ .

*Proof.* Apply Theorem 2.1 to the one-by-one expression matrix  $(f)$ . Its integer rank  $\rho$  is zero or one. A nonzero defined value forces  $\rho = 1$  by (2.5), and (2.4) then gives an invertible listed value. An invertible listed value is itself a nonzero defined evaluation. The domain guarantee also applies when  $f$  is identically zero.  $\square$

**Corollary 2.3** (Universal affine-rank tests). *Let  $n, a, b \geq 1$ ,  $s = ab(4n + 1)$ , and  $F$  have characteristic  $p$ . The list  $\mathcal{H}_{p,n,s,0}$  depends on none of the coefficients of  $L(x) = A_0 + \sum_{i=1}^n A_i x_i$ ,  $A_i \in F^{a \times b}$ , and satisfies*

$$\text{ncrk}_F L = \max_{X \in \mathcal{H}_{p,n,s,0}} \frac{\text{rank}_F(A_0 \otimes I_D + \sum_i A_i \otimes X_i^F)}{D}.$$

For a family, use  $s = (4n + 1) \sum_i a_i b_i$ ; one index attains all ranks. A rank-only oracle therefore suffices for nonadaptive testing.

*Proof.* Each affine entry has a division-free tree with at most  $4n + 1$  vertices. It remains to identify  $\tau(L)$  with free-field rank. For the homogeneous pencil  $L^{\text{hom}}(z) = \sum_{i=0}^n A_i z_i$ , pass to a sufficiently large coefficient extension. Theorem 1.5 of [17] supplies a blow-up attaining its noncommutative rank; Lemma 5.3 there shows that this rank is unchanged by field extension. The shrunk-subspace characterization bounds every normalized evaluation rank from above. Thus their supremum over extensions equals free-field rank. This use of extensions avoids the small-field regularity obstruction of [11]. Rectangular matrices are handled by zero padding. The free-field ranks of  $L$  and  $L^{\text{hom}}$  agree by coefficient-span invariance [13, Section 2, Corollary 1]. For any fixed-size witness minor of  $L^{\text{hom}}$ , multiply its nonzero polynomial by  $\det Z_0$  and pass to a purely transcendental extension. At a generic tuple both remain nonzero. Left multiplication by  $I_a \otimes Z_0^{-1}$  then produces an affine evaluation of the same rank. Conversely, affine evaluations are homogeneous ones with  $Z_0 = I$ . Thus  $\tau(L) = \text{ncrk}_F L$ , and the theorem applies.  $\square$

This extends [21, Corollary 6.1] to every coefficient field. White-box rank algorithms already exist [14, 17, 7]; a full-value oracle also recovers coefficients from  $L(0)$  and  $L(e_i) - L(0)$ . Oracle evaluation and rank computation are separate from the generator's bit bound.

**Proposition 2.4** (List lower bounds). *For  $n \geq 1$ , the simultaneous-rank guarantee implies  $L \geq \lfloor s/3 \rfloor + 1$ , even at height zero. For prime-field outputs,*

$p > 0$  and  $s \geq 4$  also imply  $n \leq p^{D^2L}$ . Unit or conjugate outputs are not required.

*Proof.* Over an algebraic closure, choose an eigenvalue  $c_j$  of  $X_1^{(j)}$ . The size-three targets  $x_1 - c_j$  each have operational rank one, but tuple  $j$  makes its assigned target singular. Their total size  $3L$  must therefore exceed  $s$ . Next, two variables with identical signatures across the list would make  $\begin{pmatrix} x_i & x_k \\ 1 & 1 \end{pmatrix}$  have rank  $D$  everywhere, although its operational rank is two. This matrix has size four. The  $n$  signatures must thus be distinct; there are at most  $p^{D^2L}$ .  $\square$

Full-rank attainment therefore needs lists already at size three, even at height zero, whereas division-free scalar nonzero detection admits single-point tests [20, 23]. At fixed  $p, s$ , variable-independent dimension forces  $L = \Omega(\log n)$ ; the polynomial gap in  $s$  remains open.

The list also tests  $r$ -vertex branching programs with affine labels at budget  $(r+1)^2(4n+1)$  and height zero. The Formal Supplement, Section S5, gives the corollary and checked triangular reduction [16, Section 6.1].

### 3. STRICT DOMAINS AND INTEGRAL MAXIMUM RANKS

Call a dimension *permitted* if it is positive and prime to the characteristic; every positive dimension is permitted in characteristic zero. A square affine pencil is *feasible* if some finite matrix substitution over an extension makes it invertible. Its evaluation is  $A(X) = A_0 \otimes I_d + \sum_i A_i \otimes X_i$ .

**Lemma 3.1** (Division lift). *Let  $K$  contain a primitive  $d$ th root  $\omega$ , with  $d$  permitted. Set  $\Omega e_j = \omega^j e_j$  and  $S e_j = e_{j+1}$ , with indices modulo  $d$ . For  $R \in \text{Mat}_d(K)$  put*

$$R^* = \sum_{a,b=0}^{d-1} c_{ab}(R) \lambda^a \mu^b \Omega^a S^b, \quad c_{ab}(R) = \frac{1}{d} \sum_{j=0}^{d-1} \omega^{-aj} R_{j,j-b}.$$

*These lifts specialize to  $R$  at  $(\lambda, \mu) = (1, 1)$  and lie in a common division algebra of degree  $d$ , represented over  $K(\lambda, \mu)$ . Every strictly defined tree remains defined after simultaneous lifting, and no nonzero minor of an evaluated expression matrix becomes zero. The rank of a rectangular matrix over this division algebra, in its matrix representation, is a multiple of  $d$ .*

*Proof.* This is the lift of [21, Lemma 3.1], now in every permitted characteristic; related rank regularity is developed in [18, 17]. Fourier inversion gives the specialization. The elements  $X = \lambda \Omega$ ,  $Y = \mu S$  satisfy  $XY = \omega YX$ ; distinct parameter monomials and a leading-term argument show that their algebra is a domain: in a product, the lexicographically largest exponent is the sum of the two largest exponents and has nonzero coefficient. Parameter monomials  $\lambda^a \mu^b$  also prove linear independence of all ordered monomials  $X^a Y^b$ . Central localization is still a domain and has dimension  $d^2$  over  $K(\lambda^d, \mu^d)$ , with basis  $X^a Y^b$  for  $0 \leq a, b < d$ . Finite dimensionality makes

every nonzero element invertible. Division-ring Gaussian elimination proves the rank assertion.

For domains use [21, Lemma 3.3] in the local ring  $K[\lambda, \mu]_{(\lambda-1, \mu-1)}$ . Induction through the actual tree preserves its specialization. At each inverse gate the operand determinant specializes to a nonzero value, hence is a unit of this local ring; its inverse agrees with the division-algebra inverse. The same specialization preserves every stipulated nonzero minor.  $\square$

**Lemma 3.2** (Permitted scalar witnesses). *An admissible scalar tree has a domain witness of permitted dimension. If it has a nonzero value, it has a permitted invertible witness.*

*Proof.* For admissible scalar trees of height at most  $h$ , write  $D_h$ , existence of a permitted domain witness; and  $I_h$ , existence of a permitted invertible witness whenever a nonzero defined value exists. Prove them in the order  $D_0, I_0, D_1, I_1, \dots$ . For  $D_0$  use scalar inputs. For  $D_h$ , multiply the operands of all outermost inverse gates. At a domain witness each factor is invertible; the product is therefore nonzero and has height at most  $h-1$ . An  $I_{h-1}$  witness makes every factor invertible, including its nested domain obligations. For  $I_h$ , start with a nonzero witness. If its dimension is divisible by the positive characteristic, add by direct sum a permitted domain witness supplied by  $D_h$  over that witness field: the same nonzero witness proves admissibility there. The new dimension is permitted and the value remains nonzero. Lemma 3.1 makes it a nonzero division-algebra element, hence a unit. This also proves  $I_0$ .  $\square$

**Lemma 3.3** (Simultaneous strict conditions). *Finitely many strict-domain and invertibility conditions, each with a permitted witness, have a common permitted witness at unit inputs.*

*Proof.* An invertibility condition requires a scalar tree value or a selected square block submatrix to be invertible on its strict domain. Repeat the witnesses to a common multiple of their dimensions, then use generic matrices of that dimension. To justify this without cancelling domains, at each fixed dimension retain a scalar polynomial  $p_T$  and polynomial matrix  $N_T$  with value  $N_T/p_T$  and strict domain exactly  $p_T \neq 0$ . At leaves  $p_T = 1$ ; the recursions are

| $T$      | $p_T$          | $N_T$                            |
|----------|----------------|----------------------------------|
| $U + V$  | $p_U p_V$      | $p_V N_U + p_U N_V$              |
| $UV$     | $p_U p_V$      | $N_U N_V$                        |
| $U^{-1}$ | $p_U \det N_U$ | $p_U^2 \operatorname{adj} N_U$ . |

For the inverse step, the domain is  $\{p_U \neq 0, \det N_U \neq 0\} = \{p_U \det N_U \neq 0\}$ , and the value is  $p_U \operatorname{adj} N_U / \det N_U = p_U^2 \operatorname{adj} N_U / (p_U \det N_U)$ . Each witness proves its condition is a nonzero polynomial. Their product, together with all input determinants, stays nonzero over  $F$ . No common original

witness field is required. Least common multiples of permitted dimensions are permitted.  $\square$

**Theorem 3.4** (Simultaneous division witnesses). *Every finite nonempty admissible matrix family over  $F$  has a common strictly defined evaluation, at unit inputs in a permitted dimension  $d$  over an extension, whose gate values lie in one division algebra. Each operational rank  $\tau(M_i)$  is an integer, attained there as  $\text{rank } M_i = d\tau(M_i)$ . If  $\tau(M_i) > 0$ , a  $\tau(M_i) \times \tau(M_i)$  block submatrix is invertible there.*

*Proof.* For one matrix  $M$ , summing zero multiples of its entries supplies a scalar tree with exactly its domain, and hence, by Lemma 3.2, a permitted domain witness of dimension  $d_0$ . Put  $t = \lceil \tau(M) \rceil$ . For  $t > 0$  choose a witness with rank ratio  $> t - 1$ . Repeat it sufficiently many times and add the  $d_0$  witness, keeping that inequality. In positive characteristic take the repetition count divisible by the characteristic, making the resulting dimension  $d'$  permitted. The two witness fields embed in a common extension: take a prime quotient of their tensor product, followed by its fraction field. After lifting, rank is a multiple of  $d'$  and at least  $td'$ , while its definition bounds it by  $\tau(M)d'$ . Thus  $\tau(M) = t$  and is attained. For  $t = 0$  lift the domain witness. Division-ring elimination gives the required block minor. Finally impose all the selected minor and domain conditions simultaneously by Lemma 3.3, include unit inputs, and lift once more.  $\square$

This existential witness dimension does not enter the quantitative bounds. Rounding uses fresh parameters; rank divisibility over small finite base fields can fail [11].

**3.1. Linear certificates.** On its strict domain, a tree  $T$  has a representation  $u_T A_T^{-1} v_T$ , with constant selectors and an affine pencil of order  $|T| + 1$ . For a leaf  $f$  use

$$A_f = \begin{pmatrix} 1 & -f \\ 0 & 1 \end{pmatrix}, \quad u_f = (1, 0), \quad v_f = (0, 1)^T.$$

Addition uses a block diagonal. Multiplication and inversion use

$$A_{UV} = \begin{pmatrix} A_U & -v_U u_V \\ 0 & A_V \end{pmatrix}, \quad A_{U^{-1}} = \begin{pmatrix} 0 & u_U \\ -v_U & A_U \end{pmatrix},$$

with the evident product selectors and the leading selector for an inverse. These are the standard operations [16, 8]; the literal selectors follow [21, Lemma 4.1]. For inverse-free trees  $A_T$  is unipotent. Addition and multiplication give block-triangular pencils, so  $\det A_T$  is the product of  $\det A_g$  over outermost inverse-rooted subtrees  $g$ . Induction and the inverse Schur complement therefore give two facts: all inverse-node  $A_g$  are invertible exactly on the strict domain, and there  $T = u_T A_T^{-1} v_T$ .

**Proposition 3.5** (Weighted certificates). *A matrix family of total size  $s$  and height at most  $h$  admits feasible affine certificates, each of order at most  $3s$ , whose order sum is at most  $3s(h+1)$ . Simultaneous invertibility of these certificates preserves every entry domain and attains all operational ranks.*

*Proof.* Extend [21, Proposition 4.2] by retaining  $A_g$  for every original inverse subtree  $g$ . Bottom-up induction and the Schur complement show that their invertibility is equivalent to the original strict domain. For each positive operational rank  $t$ , select the invertible  $t \times t$  block minor furnished by Theorem 3.4. With  $A$  the block diagonal of its entry realizations and  $UA^{-1}V$  equal to that minor, add

$$\begin{pmatrix} A & V \\ -U & 0 \end{pmatrix}.$$

On the common domain  $A$  is invertible by the realization induction, so its Schur complement is the selected minor. It is feasible at the division witness and enforces rank at least  $td$  elsewhere; the operational upper bound makes this equality. No rank certificate is needed for rank zero.

Each vertex lies in at most  $h$  such subtrees, its inverse ancestors. The domain orders sum to at most  $2hs$ . Rank certificates have total order at most  $3s$ : their selected entries are disjoint among family members, each entry size is at least one, and each  $t$  is bounded by the number of selected entries. Every individual order is at most  $3s$ , and  $2hs + 3s \leq 3s(h+1)$ .  $\square$

The selected minors and all target certificates are used only in the correctness proof. They are never inputs to the generator.

#### 4. COMPRESSING NAMES AND RATIONAL EXPRESSIONS

A substitution is *honest* if it preserves admissibility and operational rank. We establish honesty using finite matrix witnesses, retaining every original inverse obligation throughout.

**Lemma 4.1** (Support separation). *There is a deterministic polynomial-time family  $\mathcal{N}$  of maps  $[n] \rightarrow [m_*]$ , with  $m_* \leq 514s^4$ , such that every set of at most  $s$  names is injectively mapped by some member and*

$$(4.1) \quad |\mathcal{N}|\Lambda(s) \leq 16s^2\Lambda(n).$$

*Proof.* Here is the implemented weighted construction. Put  $b = 2^{\Lambda(s)}$  and choose an auxiliary prime  $16s^2b \leq \pi \leq 128s^3$  by bounded search; use a power-of-two palette  $m_* = 2^{\Lambda((4sb)^2)} \leq 514s^4$  containing these residues. Bertrand's postulate supplies  $\pi$ :  $b < 2(s+2) \leq 4s$  for  $s \geq 2$ , and 67 works for  $s = 1$ . The wider power-of-two palette is the exact implemented choice; it permits fixed-length binary names without changing the asymptotic bounds. If  $n \leq \pi$ , one injective map suffices. Otherwise label the  $n$  names distinctly in  $\mathbb{F}_\pi^m$ , with  $m = \lceil \log_\pi n \rceil$ , and enumerate its linear functionals. A pair of distinct labels collides in exactly a  $1/\pi$  fraction of them.

Give each ordered pair of distinct labels weight  $b^{c_{ij}}$ , where  $c_{ij}$  is its collision count so far. At each step select a functional of minimum colliding weight. After  $H$  selections the total weight is at most  $n^2(1 + (b - 1)/\pi)^H$ . Put  $t = \lfloor \Lambda(n)/\Lambda(s) \rfloor + 1$  and  $H = 8s^2t$ . Since  $\log_2 e < 3/2$ , the logarithm of this total is less than  $2\Lambda(n) + 3t/4 < 4t\Lambda(s)$ . Thus  $c_{ij} < 4t$ . The at most  $s^2$  ordered pairs in a support exclude fewer than  $4s^2t < H$  choices. Some selected functional separates it. Because  $n > \pi \geq s$ , we have  $t\Lambda(s) \leq 2\Lambda(n)$ , proving (4.1). The candidate table has at most  $\pi n$  rows; its entries and the weights have polynomial bit length. Exhaustive minimum selection and bounded prime search therefore have polynomial bit cost. The one-map case also satisfies (4.1).  $\square$

This is the usual wider-alphabet saving for perfect hash families [6]. Its role here is to save one logarithm in the final list bound. The auxiliary prime is unrelated to the target characteristic. An injective renaming on a used support preserves strict domains and all values; copying a palette tuple back through any map always gives exact evaluation pullback.

**Lemma 4.2** (An honest logarithmic dictionary). *Let  $g_i$  be the length- $\Lambda(m_*)$  binary spelling of  $i - 1$  in the positive letters  $u, v$ , least significant bit first, followed by  $u$ . The substitution*

$$(4.2) \quad x_i \longmapsto g_i v g_i^{-1}$$

*is honest over every coefficient field: it preserves admissibility and operational rank of expression matrices. Its literal leaf cost is exactly  $5(\Lambda(m_*) + 1) + 5$ , and it increases inversion height by at most one. On unit pairs, the substituted trees are defined and their values are mutually conjugate units; evaluation pulls back through every original gate.*

*Proof.* Call a dictionary *relatively honest* if it is honest even with any finite set of auxiliary variables left unchanged. We first prove that, for  $m, t \geq 1$ , the dictionary

$$x_{b,j} \longmapsto A^j B_b A^{-j} \quad (1 \leq b \leq m, 0 \leq j < t)$$

is relatively honest. Close a target family under all cyclic shifts  $j \mapsto j + z \pmod{t}$  and use Theorem 3.4 to obtain one unit tuple attaining the ranks of this enlarged family in dimension  $d$ . Let  $B_b$  be block diagonal with blocks  $X_{b,j}$ , let  $A$  cyclically permute the  $t$  blocks so that conjugation shifts  $j$  forward, and repeat every auxiliary input unchanged on the diagonal. Every decoded target is a direct sum of its shifted evaluations. Every original inverse is defined in every block, and its rank is  $dt$  times its original operational rank. Conversely any defined substituted tree pulls back to an original evaluation, so its normalized rank cannot increase. This proves the claim in every characteristic, with no restriction on  $t$ .

Relative honesty is preserved by composition, selecting subfamilies, and combining dictionaries on disjoint alphabets: substitute one alphabet at a

time, treating the others as auxiliary variables. We now give the induction that produces the exact short words. Let  $\mathcal{W}_0 = \{\emptyset\}$  and

$$\mathcal{W}_{N+1} = \{v^b u w : 0 \leq b \leq N, w \in \mathcal{W}_{N-b}\}.$$

These are exactly the positive words of length  $N + 1$  ending in  $u$ . Define group expressions on  $z_0, \dots, z_N$  by

$$\begin{aligned} T_{0,\emptyset} &= z_0, \\ T_{N+1,v^b u w}(z_0, \dots, z_{N+1}) &= z_0^b T_{N-b,w}(z_1, \dots, z_{N-b+1}) z_0^{-b}. \end{aligned}$$

This whole dictionary is relatively honest by induction. Indeed, first form the branch dictionaries on disjoint alphabets  $z_{b,j}$ . Then apply the cyclic dictionary  $z_{b,j} \mapsto z_0^b z_{j+1} z_0^{-b}$ , using a grid of  $(N+1)^2$  variables and selecting the required ones. Simultaneous conjugation commutes with every group expression, giving exactly the displayed recurrence.

Finally substitute  $z_j = u^j v u^{-j}$ , another relatively honest cyclic dictionary. On unit inputs the recurrence gives  $T_{N,w} = w v w^{-1}$ . Select the words  $g_i \in \mathcal{W}_{\Lambda(m_*)+1}$ . For any admissible target family, the composite coded family is admissible with the same ranks. Apply Theorem 3.4 to this family to obtain a unit input attaining all these ranks. At that input both the composite and literal dictionaries are strictly defined and agree. Induction through every original gate transfers their values and inverse certificates to (4.2). Conversely, every defined literal substitution pulls back to a target evaluation, so its normalized rank cannot increase. This proves honesty without identifying the two trees' domains at arbitrary inputs. Spell  $g_i^{-1}$  in reverse order with one inverse gate per letter, and retain a terminal 1 in each word product. Each letter and its inverse contribute three nodes; the two word products contribute two multiplication nodes per letter. The terminal leaves, central  $v$  and two outer products add five, giving exactly  $5|g_i| + 5$  nodes and height one, without cancellation.  $\square$

This is the finite-witness proof in the Lean development; auxiliary variables are essential to its branch induction. It uses no free-field embedding or fixed-dimension domain-equivalence assumption.

## 5. A UNIVERSAL QUOTIENT DESIGN

We prove the design over an arbitrary field and after every scalar extension. It converts feasibility into a rank within less than one block of full rank. This quantitative gap is what the division lift will turn into invertibility.

**Lemma 5.1** (Overlap on a tree). *Let  $\sum_{i=1}^w A_i x_i$  be a feasible homogeneous pencil of order  $a$  over  $F$ , let  $E/F$  be an extension, and let  $\mathcal{T}$  be a tree on  $[w]$ . For arbitrary  $P_i \in E^{d_0 \times a}$  and  $t = \text{rank}[P_1 \cdots P_w]$ ,*

$$(5.1) \quad \sum_i \text{rank } P_i + \text{rank} \left( \sum_i P_i A_i \right) + \sum_{ij \in \mathcal{T}} \text{rank}(P_i - P_j) \geq 2t.$$

*Proof.* Put  $S_i = \text{col } P_i$ ,  $R_i = \text{row } P_i$ ,  $\Delta = \sum_i \dim S_i$ , and  $\epsilon = \Delta - t$ . Pass to a common extension of  $E$  and an invertible witness field; ranks do not change under scalar extension. At an invertible witness  $\sum_i A_i \otimes Y_i$  of dimension  $b$ , the block matrix with first column  $A_i \otimes Y_i$  and remaining columns the signed tree incidence matrix tensored with  $I_{ab}$  is invertible. Indeed, the incidence columns span the kernel of block-row summation, and the first column is invertible on the quotient. Act on row vectors from the right and apply this matrix to  $\bigoplus_i (R_i \otimes E^b)$  and compare image dimensions to get

$$\dim \sum_i R_i A_i + \sum_{ij \in \mathcal{T}} \dim(R_i + R_j) \geq \Delta.$$

Factor  $P_i = C_i Q_i$  using column and row bases. Sylvester's rank inequality, applied to  $[C_1 \cdots C_w](Q_i A_i)_i$  and to  $[C_i \ C_j](Q_i; -Q_j)$ , gives

$$\begin{aligned} \text{rank} \sum_i P_i A_i &\geq \dim \sum_i R_i A_i - \epsilon, \\ \text{rank}(P_i - P_j) &\geq \dim(R_i + R_j) - \dim(S_i \cap S_j). \end{aligned}$$

Adding vertices in rooted-tree order shows  $\sum_{ij \in \mathcal{T}} \dim(S_i \cap S_j) \leq \epsilon$ . The last two terms in (5.1) therefore sum to at least  $\Delta - 2\epsilon$ ; adding  $\Delta$  proves the result.  $\square$

**Lemma 5.2** (Full-window folded multiplicity). *Let  $S \subset E[z]_{<d_0}$  have dimension  $1 \leq t \leq r$ . Suppose  $\gamma$  has order  $> d_0 - 1$  and the extended windows  $\{\alpha_j \gamma^v : -t + 1 \leq v \leq r - 1\}$  are pairwise disjoint, each has  $r + t - 1$  distinct points, and they avoid zero. Let  $Z_j$  be the dimension of the subspace of  $S$  vanishing on  $\alpha_j, \dots, \alpha_j \gamma^{r-1}$ . Then*

$$(5.2) \quad \sum_j Z_j(r - t + Z_j) \leq t(d_0 - t).$$

*Proof.* For a basis  $f_1, \dots, f_t$  use the folded determinant

$$C(z) = \det(f_i(\gamma^b z))_{0 \leq b < t, 1 \leq i \leq t}.$$

Distinct leading degrees give a nonzero Vandermonde leading coefficient and degree at most  $t(d_0 - 1) - t(t - 1)/2$ . Constant basis changes multiply  $C$  by a nonzero scalar and preserve multiplicities. A basis with distinct vanishing orders at zero gives the factor  $z^{t(t-1)/2}$ . The remaining degree is at most  $t(d_0 - t)$ . At  $z = \alpha_j \gamma^v$  put  $L_v = |[0, t - 1] \cap [-v, r - 1 - v]|$ . Choosing the first  $Z_j$  basis members to vanish on the window gives a zero  $L_v \times Z_j$  submatrix and hence rank defect at least  $\max(0, Z_j + L_v - t)$ . A rank defect  $b$  forces determinant vanishing of order at least  $b$ , by constant row operations. Interior shifts contribute  $Z_j(r - t + 1)$  and each boundary contributes  $\sum_{b=1}^{Z_j-1} (Z_j - b) = Z_j(Z_j - 1)/2$ . Adding these disjoint root multiplicities proves (5.2) in every characteristic.  $\square$

The folded determinant is classical [15, 12]. The boundary count above makes explicit the additional roots discussed in [5, ECCC version, Remark 6.4].

**Proposition 5.3** (Constant rank gap). *Let  $q \geq 1$ ,  $w \geq 2$ , and set*

$$\begin{aligned} \eta &= qw, & \nu &= (w-2)q+1, & \delta &= w\nu, \\ r &> q\delta, & d_0 &= 2r-\nu, & \ell &= r+\eta-1. \end{aligned}$$

*Given a scalar  $\gamma$  of order at least  $2w\ell$ , one can compute  $R_1, \dots, R_w \in \text{Mat}_r(K)$  so that, for every coefficient extension and every feasible order- $a$  pencil with  $1 \leq a \leq q$ ,*

$$(5.3) \quad \text{rank} \sum_i A_i \otimes R_i > (a-1)r.$$

*The construction uses polynomially many field operations.*

*Proof.* Put  $\alpha_j = \gamma^{(j-1)\ell+\eta-1}$  for  $1 \leq j \leq 2w$  and  $V_j = ((\alpha_j \gamma^b)^u)_{0 \leq u < d_0, 0 \leq b < r}$ . Let  $U_i$  insert  $V_1$  in row block  $i$  of  $K^{wd_0 \times r}$ . Form  $C_0 \in K^{wd_0 \times (2w-1)r}$  from one signed  $V_j$  column for each edge  $\{1, j\}$  of the star,  $2 \leq j \leq w$ , with signs  $+$  at row block 1 and  $-$  at row block  $j$ , and a private  $V_{w+i}$  in each row block  $i$ .

The augmented matrix

$$[\sum_i A_i \otimes U_i \mid I_a \otimes C_0]$$

has full row rank. Otherwise write a nonzero left null vector as blocks  $y_{k,i} \in E^{1 \times d_0}$  and define column  $k$  of  $P_i$  to be  $y_{k,i}^\top$ . The polynomial coefficient matrices  $P_i \in E^{d_0 \times a}$  then satisfy  $1 \leq t = \text{rank}[P_1 \cdots P_w] \leq aw \leq \eta < r$ . At the first window  $\sum_i P_i A_i$  vanishes, at edge windows the respective differences vanish, and at private windows  $P_i$  vanishes. Lemma 5.1 gives  $Z = \sum_j Z_j \geq 2t$  for the space spanned by all their columns. But Cauchy's inequality gives

$$\begin{aligned} \sum_j Z_j(r-t+Z_j) &\geq (r-t)Z + \frac{Z^2}{2w} \\ &\geq 2t(r-t) + \frac{2t^2}{w} \geq t(d_0-t+1), \end{aligned}$$

contradicting Lemma 5.2. Its extended-window exponents lie in  $[(j-1)\ell, j\ell-1]$ , because  $t \leq \eta$ ; the order bound on  $\gamma$  gives the required joint distinctness. The last inequality follows from  $(w-2)t/w \leq \nu-1$ . These are inequalities of integer dimensions, not divisions inside the field.

Compute a full-row-rank  $H$  with  $\ker H = \text{col } C_0$ , and let  $c = wd_0 - \text{rank } C_0$ . Column counting gives  $c \geq r - \delta$ . The augmented rank assertion gives  $\text{rank} \sum_i A_i \otimes HU_i = ac$ ; the order-one pencil  $x_1$  also gives  $c \leq r$ . Pad  $HU_i$  with zero rows to make  $R_i$  square. Then  $ac \geq a(r-\delta) > (a-1)r$ . The computed full-kernel identity survives every scalar extension.  $\square$

## 6. A DETERMINANT CURVE AND DESCENT

Fix the three-variable case  $w = 3$  of Proposition 5.3, with  $r$  permitted and a primitive  $r$ th root  $\zeta$ . In each geometric component, apply Lemma 3.1 over  $\overline{F}(\lambda, \mu)$ : specialization preserves a rank above  $(a-1)r$ , and divisibility

by  $r$  makes it  $ar$ . Thus the lift makes every feasible pencil invertible. A denominator-free version of the lift will make both sampling and bit bounds transparent.

For  $R \in \text{Mat}_r(K)$  define

$$(6.1) \quad \begin{aligned} h_b(T) &= \sum_{j=0}^{r-1} \zeta^j R_{j,j-b} \prod_{0 \leq v < r, v \neq j} (T - \zeta^v), \\ \tilde{R} &= \sum_{b=0}^{r-1} h_b(X) Y^b. \end{aligned}$$

Lagrange interpolation at the roots of  $T^r - 1$ , whose derivative there is  $r\zeta^{-j}$ , gives  $\tilde{R} = rR^*$ . Thus all feasible pencil determinants remain nonzero. Each parameter has degree at most  $r - 1$  in a lifted entry. Conjugation by clock and shift scales the two parameters by  $r$ th roots:  $S^{-1}\Omega S = \zeta\Omega$  and  $\Omega^{-1}S\Omega = \zeta^{-1}S$ . Determinants are unchanged by these similarities, so their nonzero monomials have both exponents divisible by  $r$ . Every order- $a$  pencil determinant has the form

$$P(U, V), \quad U = \lambda^r, \quad V = \mu^r, \quad \deg_U P, \deg_V P \leq a(r - 1).$$

**Lemma 6.1** (Direct central-curve realization). *Set  $k_0 = q(r - 1) + 1$  and  $\kappa = (r - 1)(q(r - 1) + 2)$ . Define*

$$X_T e_j = \begin{cases} e_{j+1} & j < r - 1, \\ T e_0 & j = r - 1, \end{cases} \quad J e_j = \zeta^{-j} e_j, \quad Y_T = X_T^{k_0} J.$$

*Substituting  $(X_T, Y_T)$  in (6.1) makes every feasible order- $a$  determinant a nonzero polynomial in  $T$  of degree at most  $a\kappa$ . Evaluation uses polynomially many arithmetic operations and no root extraction.*

*Proof.* Direct computation gives

$$X_T Y_T = \zeta Y_T X_T, \quad X_T^r = T I, \quad Y_T^r = \varepsilon T^{k_0} I, \quad \varepsilon = (-1)^{k_0(r-1)}.$$

Over an algebraic closure of  $K(T)$ , the pair is simultaneously conjugate to  $(\lambda\Omega, \mu S)$  with  $\lambda^r = T$ ,  $\mu^r = \varepsilon T^{k_0}$ : use an  $X_T$  eigenvector and its successive  $Y_T$  translates. Indeed,  $X_T$  has distinct eigenvalues since  $T \neq 0$  and the characteristic does not divide  $r$ ; the relation  $X_T Y_T = \zeta Y_T X_T$  cycles their eigenlines. Rescaling this cyclic basis makes every transition equal to a chosen  $\mu$ . The identity  $(X_T^{k_0} J)^m = \zeta^{-k_0 m(m-1)/2} X_T^{k_0 m} J^m$  proves the stated sign at  $m = r$ . Hence the determinant is exactly  $P(T, \varepsilon T^{k_0})$ . The exponents  $1, k_0$  inject the box  $0 \leq \deg_U P \leq q(r - 1)$  into distinct powers, so this polynomial is nonzero and has degree at most  $a(r - 1)(k_0 + 1) = a\kappa$ . The sign is necessary when  $r$  is even. For implementation, a monomial acts by

$$X_T^a Y_T^b e_j = \zeta^{-k_0 b(b-1)/2 - bj} T^{\lfloor (j+a+k_0 b)/r \rfloor} e_{(j+a+k_0 b) \bmod r}.$$

This uses bounded powers and sums, with no sample-dependent denominator.  $\square$

**Proposition 6.2** (Simultaneous sampling over every coefficient field). *Let  $A/K_p$  be finite étale of dimension  $e$ , and suppose the design nodes retain their required distinctness and the root has order  $r$  under each embedding  $A \rightarrow \overline{K_p}$ . A family of feasible homogeneous pencils with maximum order  $q$  and total order at most  $W$  is simultaneously invertible, after regular representation over  $K_p$ , at some one of any  $eW\kappa + 1$  nonzero samples whose images are distinct under each such embedding.*

*Proof.* Fix the target coefficient field  $F$ . After extending to an algebraic closure,  $A \otimes_{K_p} \overline{F} \cong \overline{F}^e$ . In each component, keeping target coefficients fixed, every target determinant is a nonzero polynomial of degree at most its order times  $\kappa$ . Summing these root bounds over all targets and all  $e$  components excludes at most  $eW\kappa$  sample indices. At a remaining index every component is invertible. Regular representation splits as their direct sum, so the descended pencils are invertible over  $F$  as well. This argument requires neither irreducibility of  $A$  nor linear disjointness from  $F$ .  $\square$

As in [21, Sections 3 and 5], for affine pencils in  $u, v$ , homogenize with a third variable  $x_0$  and include the normalizer pencil  $x_0$ . Write the descended sample as  $(Z_0, Z_1, Z_2)$ . Discard samples with singular  $Z_0$ , and set  $U = Z_0^{-1}Z_1$ ,  $V = Z_0^{-1}Z_2$ . The identity

$$A_0 \otimes I + A_1 \otimes U + A_2 \otimes V = (I \otimes Z_0^{-1}) \sum_{i=0}^2 A_i \otimes Z_i$$

preserves every successful sample. Include the unit conditions on  $u, v$  among the target certificates, discard any remaining nonunit pairs, and decode (4.2). This filtering depends only on the generated matrices, not on any target. A good sample survives both guards, and every printed decoded coordinate is a conjugate of  $V$ . All original inverse gates are enforced by their certificates.

## 7. UNIFORM ARITHMETIC AND BIT BOUNDS

Regular representation uses increasing power bases of monic quotients. Each rebuilt field has its own monic power basis; matrix indices are flattened in row-major order, as specified in the artifact. The construction must work with binary-encoded prime characteristic. Consequently an enumeration of all  $p$  residues is legitimate only when  $p$  is polynomially bounded in the unary parameters. The following choices make this distinction explicit. In all branches arithmetic is chosen using the capacity  $W_{\max} = 4q^2$ ; the actual weight  $W \leq W_{\max}$  affects only the sample count.

Characteristic two. Use the conservative, fully verified parameters

$$m_2 = 2\Lambda(q) + 5, \quad r = 2^{m_2} - 1, \quad e = 4m_2, \quad A = K = \mathbb{F}_{2^e}.$$

They satisfy

$$31q^2 < r \leq 512q^2, \quad e \leq 28\Lambda(q), \quad 2^e = (r+1)^4, \quad eW\kappa + 1 \leq 2^e - 1.$$

Also  $r > 3q(q+1)$  and  $6(r+3q-1) < 2^e - 1$ . An irreducible defining polynomial and a multiplicative generator  $\gamma$  can therefore be found by exhaustive finite-field searches: the entire field has cardinality at most  $513^4 q^8$ . Use  $\zeta = \gamma^{(2^e-1)/r}$ , the powers of  $\gamma$  for design nodes, and  $eW\kappa + 1$  distinct nonzero elements for samples. The resulting bounds are

$$(7.1) \quad D = er \leq 14336q^2\Lambda(q), \quad L_{\text{core}} \leq 7340032Wq^5\Lambda(q) + 1.$$

The inequalities include  $q = 1$ ; no exceptional input is suppressed.

Odd characteristic. Choose the least power of two  $r > 3q(q+1)$ , so  $r \leq 6q(q+1) \leq 12q^2$ . Put  $N_0 = r/2$ ,  $M = 6(r+3q-1)$ ,  $B = \binom{M}{2}$ , and  $T_* = 4q^2\kappa + B + 1$ . Set

$$E_0 = 2\lceil \log_2(2(T_* + 1)) \rceil, \quad E = \max(N_0, 2E_0), \quad P = ET_* + 2.$$

If  $p > P$ , use the intact algebra  $A = \mathbb{F}_p[z]/(z^{N_0} + 1)$ , of dimension  $N_0$ , and construct the design over  $K = \mathbb{F}_p$ . It is étale, and every component root has order  $r$ . This branch requires no factorization of either  $p-1$  or the cyclotomic polynomial.

If  $p \leq P$ , obtain an irreducible factor of  $z^{N_0} + 1$  by Berlekamp splitting [4]. Compute the Frobenius fixed space; a nonconstant fixed element and its gcds after subtracting  $c \in \mathbb{F}_p$  split any reducible squarefree factor. Here enumeration of  $p$  scalars is polynomially bounded. Starting with the resulting field, while  $Q = p^e$  fails  $Q - 1 > eT_* + 1$ , rebuild a field of degree  $2e$  over  $\mathbb{F}_p$  by a bounded search for a monic irreducible polynomial. Reselect an  $r$ th primitive root;  $r \mid p^{2e} - 1$ . Every failing stage has  $e < E_0$ : at  $E_0$ ,  $2^{E_0} \geq 4(T_* + 1)^2 > E_0(T_* + 1) + 1$ . The inequality  $2^e - 1 > e(T_* + 1)$  persists for every  $e \geq E_0$ , since  $2^{e+1} - 1 = 2(2^e - 1) + 1 > 2e(T_* + 1) \geq (e+1)(T_* + 1)$ . As  $e \geq E_0 \geq 2$ , this implies  $p^e - 1 > eT_* + 1$ . Thus the final degree is at most  $E$ . At a failing stage,  $p^{2e} \leq (eT_* + 2)^2 \leq P^2$ , so the full search and enumeration in each rebuilt field are polynomially bounded.

In either branch find an element of order at least  $M$  among the first  $B+1$  nonzero elements of  $K$ . At most  $B$  elements are roots of  $\prod_{i=1}^{M-1} (z^i - 1)$ . Testing these powers gives the design seed in polynomial time. Choose the first  $eW\kappa + 1$  distinct nonzero samples in  $K$ . They remain distinct under every embedding; in the large-prime branch they are scalars in every component. Since  $e = O(q^2)$ , this yields  $D = O(q^4)$  and  $L_{\text{core}} = O(Wq^7)$ . In particular the implemented branch satisfies

$$(7.2) \quad D \leq 3888q^4, \quad L_{\text{core}} \leq 314928Wq^7 + 1.$$

Quotients are always computed over the field  $K$ , never by field elimination over a reducible algebra.

Characteristic zero. Use the cyclotomic descent of [21, Section 5]: the same power-of-two order  $r$  and the algebra  $A = \mathbb{Q}[z]/(z^{r/2} + 1)$ . Set  $e = r/2$  and construct the design with  $\gamma = 2$ . Use the integer samples  $1, \dots, eW\kappa + 1$ . The bounds are

$$(7.3) \quad D \leq 162q^4, \quad L_{\text{core}} \leq 2916Wq^7 + 1.$$

**Lemma 7.1** (Stagewise rational heights). *The rational construction has design height  $O(r^2)$ , cleared quotient height  $O(r^3)$  and raw sample height  $H_{\text{raw}}$  from (7.4). The retained normalized coordinates and their inverses have a common denominator height  $b = O(D(H_{\text{raw}} + \log(D + 1)))$ . A conjugator of length at most  $k_{\text{word}}$  gives decoded height at most  $b_{\text{dec}} = 2(k_{\text{word}} + 1)(b + \Lambda(D))$ . Here height  $b$  means an integer numerator matrix and a common nonzero integer denominator whose absolute values are at most  $2^b$ .*

*Proof.* The design entries have bit height  $O(r^2)$ . If  $g$  is a nonzero pivot-minor determinant of  $C_0$ , Cramer's rule gives a full quotient with common denominator  $g$ ; Hadamard's inequality bounds  $g$  and its cleared numerators by  $O(r^3)$  bits. Multiplying the quotient by  $g$  makes its matrices integral without changing any rank gap. In the power basis of  $A$ , coefficient  $\ell_1$  norm is submultiplicative because  $z^{r/2} = -1$ . Cofactors in (6.1) add only  $O(r + \log(r + 1))$  bits. For  $J_* = eW\kappa + 1$ , all raw sample entries therefore have height

$$(7.4) \quad H_{\text{raw}} = O(r^3 + \kappa \log(J_* + 1) + \log(r + 1)).$$

Regular representation uses signed copies of these coefficients. For integer raw matrices with entries bounded by  $2^H$ , Cramer's rule gives common height  $b = D(H + \Lambda(D)) + 1$  for  $Z_0^{-1}Z_i$ . Its inverse is  $Z_i^{-1}Z_0$ , so the same bound applies after filtering. Multiplication adds the two heights and  $\Lambda(D)$ ; applying this to the conjugator and its inverse proves the decoded bound. This also bounds intermediate products. Euclidean gcd gives reduced fractions.  $\square$

In particular the rational output has at most

$$2\Lambda(D) + 2\Lambda(L) + 2 + LnD^2(4(b_{\text{dec}} + 1) + 3)$$

bits, including its header. The supplement gives the exact checked height and encoding declarations; this output bound does not replace the machine-runtime bound.

The characteristic-two saving comes from  $r = 2^{m_2} - 1$  dividing  $2^{4m_2} - 1$ , so  $e = 4m_2 = O(\log q)$ . The other uniform branches use  $e = O(q^2)$ ; the intact large-prime algebra avoids factoring  $p - 1$  and the cyclotomic polynomial. Smaller splitting degrees may improve particular inputs, but are not required for the theorem.

All branches use fixed bounded loops, computed pivots, exact polynomial quotients and matrix arithmetic. Their dimensions, iteration counts, and intermediate encodings are polynomially bounded in  $q + W + \Lambda(p) + 1$ . Section 9 describes the checked compilation of these operations to actual Turing-machine transitions.

## 8. ASSEMBLY AND THE QUANTITATIVE THEOREM

We now specify the common budgets, which are independent of every target family. Let  $m_*$  be the palette in Lemma 4.1 and put

$$(8.1) \quad q = 300s\Lambda(s), \quad W = q(h+2) + 1, \quad W_{\max} = 4q^2.$$

The exact dimension is  $4m_2(2^{m_2} - 1)$  in characteristic two,  $er$  in odd characteristic, and  $r^2/2$  in characteristic zero. The list has at most  $|\mathcal{N}|(eW\kappa + 1)$  members before filtering. The following parameter flow records the dependence relevant to uniformity:

$$\begin{aligned} s &\mapsto q \mapsto (r, e) \mapsto D = er, \\ (s, h) &\mapsto W, \quad (n, s) \mapsto \mathcal{N}, \quad L \leq |\mathcal{N}|(eW\kappa + 1). \end{aligned}$$

Here  $w = 3$ ,  $\eta = 3q$ ,  $\nu = q + 1$ ,  $d_0 = 2r - \nu$ ,  $\ell = r + \eta - 1$ ,  $k_0 = q(r - 1) + 1$ , and  $\kappa = (r - 1)(k_0 + 1)$ ; Section 7 specifies  $e$  in each branch. The dictionary's literal leaf cost is at most  $96\Lambda(s)$ . Coding a renamed family therefore gives total size at most  $96s\Lambda(s)$  and height at most  $h + 1$ . Include the two scalar unit conditions  $u, v$ , charging four additional vertices conservatively. Proposition 3.5 gives maximum certificate order at most  $3(96s\Lambda(s) + 4) \leq q$  and total order at most  $q(h + 2)$ . Homogenization adds the order-one normalizer, accounted for in  $W$ . For  $h \leq s$  the displayed  $W \leq 4q^2$ .

*Proof of Theorem 2.1.* Compute the arithmetic, quotient and sample list from  $p, q, W$ , using  $W_{\max}$  to fix every field choice. Normalize, filter unit pairs, and decode to the palette. For each naming map in  $\mathcal{N}$ , copy palette coordinates to the  $n$  output positions. The algorithm contains no target-dependent step.

Fix any coefficient field and admissible family after constructing the list. (1) *Ranks and coding.* Theorem 3.4 gives integral optimal ranks and universal upper bounds. The support has at most  $s$  names; choose an injective naming map. Lemma 4.2 preserves admissibility and these ranks.

(2) *Certificates and sampling.* Proposition 3.5 gives the coded family's domain and rank certificates. Together with the unit guards and normalizer they fit (8.1); Proposition 6.2 supplies one sample satisfying all of them.

(3) *Decoding.* This sample survives the guards and pulls back to one list member with every original inverse defined. Its rank lower bounds equal the universal upper bounds from (1), giving the required equalities. Applying the same argument to the constant formula 1 proves nonemptiness. Every retained tuple consists of conjugates of one unit, a property preserved by coordinate copying.

(4) *Bounds and implementation.* The dimension is  $er$  and depends only on  $p, q$ , hence on  $p, s$ . Equations (7.1)–(7.3) and  $W = O((h + 1)q)$  give core list lengths  $O((h + 1)q^6\Lambda(q))$  in characteristic two and  $O((h + 1)q^8)$  otherwise. Since  $q = O(s\Lambda(s))$  and  $\Lambda(q) = O(\Lambda(s))$ , multiplying by  $|\mathcal{N}| = O(s^2\Lambda(n)/\Lambda(s))$  gives precisely (2.3). This accounts for the saved logarithm in both list bounds.

The arithmetic bounds, polynomial-size naming table, word decoding, and complete dense serialization give polynomial bit complexity in  $n+s+\Lambda(p)+1$ . The verified fixed-machine composition below realizes this same contract. Enlarge one absolute positive  $C$  and one positive exponent  $k$  to cover all branches and all positive inputs. This proves (2.2).  $\square$

## 9. THE VERIFIED MACHINE AND ITS EXACT CONTRACT

Already at  $s = 1$ , the selected dimensions are  $D(2, 1) = 3,355,443,100$  and  $D(0, 1) = 2,199,023,255,552$ ; `scripts/parameters.py` reproduces them. These conservative construction sizes are far from practical.

**9.1. Literal semantics and encoding.** The four-symbol machine convention and self-delimiting rational records follow OpenAI’s `RationalHitting` statement [22]; the characteristic tag and uniform prime-field output extend that contract. **Literal semantics.** Lean uses Section 2’s trees, vertex count and individual admissibility. Each inverse carries both inverse identities, and constants act through  $F \hookrightarrow E$ . Common domain witnesses are proved.

**Quantifier order.** Theorem 2.1 chooses  $G, D, C, k$  before the inputs  $p, n, s, h$ . The output precedes the coefficient field and target family; one index and its ranks then work for every  $E, d, Z$  in (2.5). The family has one total size budget; neither it nor its field enters  $G$ .

**Literal output and cost.** Fix the following self-delimiting encoding. For  $t \in \mathbb{N}$ , let  $b = 1$  at  $t = 0$  and  $b = \lfloor \log_2 t \rfloor + 1$  otherwise; encode  $t$  as  $1^b$  followed by its  $b$  binary digits, most significant first. Encode a rational number by a sign bit (0 for nonnegative, 1 for negative), the code of its numerator magnitude, and the code of its positive denominator, in reduced form. Encode a residue modulo  $p > 0$  by the natural code of its canonical representative in  $\{0, \dots, p-1\}$ . The output starts with the codes of  $D$  and  $L$ , followed by all tuples in list order, all coordinates in variable order, and every matrix entry in row-major order. The parameter  $n$  and characteristic tag are known from the input. Input is the code of  $p$  followed by  $1^n 0 1^s 0 1^h 0$ . The time bound counts ordinary deterministic bit-machine steps, including printing this entire dense output.

**Cost model.** The bit bound covers printing, not oracle evaluation or arithmetic on the unrestricted target coefficients.

**9.2. Formal correspondence.** Lean proves `TC10.MainStatement` unconditionally; `PaperContract.main` expands its quantifiers, and `validation/contract.json` pins the theorem and encoding. Verified parsing, transitions and dense printing use no oracles; the final tape contains exactly the output, from the head rightward, followed by blanks.

The Formal Supplement provides source excerpts, correspondence and reproduction details. Lean checks the star overlap, lower bounds and branching-program reduction; the general-tree strengthening and cited free-field interpretation are not separately formalized. The project pins

Lean 4.34.1 [9] and mathlib [19]; `scripts/verify.py` rebuilds and audits all proofs, allowing only `propext`, `Classical.choice` and `Quot.sound`. Optimality and necessity of the  $h + 1$  factor remain open.

#### ACKNOWLEDGMENTS

The preparation of this manuscript and its computational companion benefited from assistance by AI systems developed by OpenAI and Anthropic. Responsibility for the mathematical claims, implementation, and presentation rests with the author.

#### REFERENCES

- [1] S. A. Amitsur. Rational identities and applications to algebra and geometry. *Journal of Algebra*, 3:304–359, 1966. doi:10.1016/0021-8693(66)90004-4.
- [2] V. Arvind, A. Chatterjee, and P. Mukhopadhyay. Black-box identity testing of noncommutative rational formulas in deterministic quasipolynomial time. *SIAM Journal on Computing*, pp. STOC24-260–STOC24-296, 2026 (online 1 June). doi:10.1137/24M1689508. Full version arXiv:2309.15647v4, 2025.
- [3] V. Arvind and P. S. Joglekar. A multivariate to bivariate reduction for noncommutative rank and related results. In *ICALP, LIPIcs* 297, 14:1–14:19, 2024. doi:10.4230/LIPIcs.ICALP.2024.14.
- [4] E. R. Berlekamp. Factoring polynomials over finite fields. *Bell System Technical Journal*, 46:1853–1859, 1967. doi:10.1002/j.1538-7305.1967.tb03174.x.
- [5] J. Brakensiek, Y. Chen, M. Dhar, and Z. Zhang. From random to explicit via subspace designs with applications to local properties and matroids. *ECCC TR25-146*, 2025; arXiv:2510.13777. Accepted at *STOC 2026*. Remark 6.4 is cited from the ECCC version.
- [6] N. H. Bshouty. Linear time constructions of some  $d$ -restriction problems. In *CIAC, LNCS* 9079, 74–88, 2015. doi:10.1007/978-3-319-18173-8\_5.
- [7] A. Chatterjee, S. Ghosh, R. Gurjar, R. Raj, and T. Thierauf. Bipartite matching is in NC. *ECCC TR26-100*, revision 2, 15 July 2026. Pinned manuscript, Section 5.
- [8] P. M. Cohn. *Free Ideal Rings and Localization in General Rings*. Cambridge University Press, 2006. doi:10.1017/CBO9780511542794.
- [9] L. de Moura and S. Ullrich. The Lean 4 theorem prover and programming language. In *CADE 28, LNCS* 12699, 625–635, 2021. doi:10.1007/978-3-030-79876-5\_37.
- [10] H. Derksen and V. Makam. Polynomial degree bounds for matrix semi-invariants. *Advances in Mathematics*, 310:44–63, 2017. doi:10.1016/j.aim.2017.01.018.
- [11] H. Derksen and V. Makam. The regularity lemma is false over small fields. *Communications in Algebra*, 48:1333–1339, 2020. doi:10.1080/00927872.2019.1684507.
- [12] M. A. Forbes and V. Guruswami. Dimension expanders via rank condensers. In *APPROX/RANDOM, LIPIcs* 40, 800–814, 2015. doi:10.4230/LIPIcs.APPROX-RANDOM.2015.800.
- [13] M. Fortin and C. Reutenauer. Commutative/noncommutative rank of linear matrices and subspaces of matrices of low rank. *Séminaire Lotharingien de Combinatoire*, 52:B52f, 2004. Published article.
- [14] A. Garg, L. Gurvits, R. Oliveira, and A. Wigderson. Operator scaling: theory and applications. *Foundations of Computational Mathematics*, 20:223–290, 2020. doi:10.1007/s10208-019-09417-z.
- [15] V. Guruswami and S. Kopparty. Explicit subspace designs. *Combinatorica*, 36:161–185, 2016. doi:10.1007/s00493-014-3169-1.

- [16] P. Hrubeš and A. Wigderson. Non-commutative arithmetic circuits with division. *Theory of Computing*, 11:357–393, 2015. doi:[10.4086/toc.2015.v011a014](https://doi.org/10.4086/toc.2015.v011a014).
- [17] G. Ivanyos, Y. Qiao, and K. V. Subrahmanyam. Constructive non-commutative rank computation is in deterministic polynomial time. *Computational Complexity*, 27:561–593, 2018. doi:[10.1007/s00037-018-0165-7](https://doi.org/10.1007/s00037-018-0165-7).
- [18] G. Ivanyos, Y. Qiao, and K. V. Subrahmanyam. Non-commutative Edmonds’ problem and matrix semi-invariants. *Computational Complexity*, 26:717–763, 2017. doi:[10.1007/s00037-016-0143-x](https://doi.org/10.1007/s00037-016-0143-x).
- [19] The mathlib Community. The Lean mathematical library. In *Certified Programs and Proofs (CPP)*, 367–381, 2020. doi:[10.1145/3372885.3373824](https://doi.org/10.1145/3372885.3373824).
- [20] OpenAI. One rational matrix hitting point for noncommutative formulas. 24 September 2026. OpenAI Math Release preprint, repository revision [fd4aeeb2](#). [Pinned preprint](#); formalization scope in [22].
- [21] OpenAI. Polynomial hitting lists for noncommutative rational formulas. 24 September 2026. [Pinned manuscript in the OpenAI mathematics repository](#).
- [22] OpenAI. Uniform black-box noncommutative identity testing across characteristics. Formalization scope document, result 116. [Repository revision fd4aeeb2ee4fc729c18d98444fed42fd0529eeeb](#), consulted 18 October 2026.
- [23] OpenAI. Uniform matrix hitting points in every positive characteristic. 4 October 2026. [Pinned manuscript in the OpenAI mathematics repository](#).