

# On Symmetric Rank Decompositions of the 3x3 Matrix Multiplication Tensor, Part 2

Jason Yang\* and Virginia Vassilevska Williams†

March 2023

## Abstract

This paper is an improved attempt to find a rank  $\leq 23 \bmod 2$  decomposition of the tensor describing  $3 \times 3$  matrix multiplication. We investigate a much bigger set of possible symmetry restrictions and are more systematic about what kinds of restrictions we test. Our results rule out the existence of a rank  $\leq 23$  decomposition for many more symmetry restrictions.

## 1 Introduction

Fast matrix multiplication essentially consists of finding a low-rank decomposition of a  $\langle n, k, m \rangle$  *matrix multiplication tensor*  $\mathcal{T}^{\langle n, k, m \rangle}$ , a  $n \times k \times k \times m \times m \times n$  array where

$$\mathcal{T}_{i,j,j,k,k,i}^{\langle n, k, m \rangle} = 1 \quad \forall i, j, k, \text{ and all other elements are } 0.$$

A  $R$ -rank decomposition over a field  $\mathbb{F}$  is a (multi)set of matrix triplets  $D = \{(A^{(r)} \in \mathbb{F}^{n \times k}, B^{(r)} \in \mathbb{F}^{k \times m}, C^{(r)} \in \mathbb{F}^{m \times n})\}_{0 \leq r < R}$  s.t.

$$\mathcal{T}^{\langle n, k, m \rangle} = \sum_{r=0}^{R-1} A^{(r)} \times B^{(r)} \times C^{(r)} \quad (\text{equivalently, } \mathcal{T}_{a,b,c,d,e,f}^{\langle n, k, m \rangle} = \sum_{r=0}^{R-1} A_{a,b}^{(r)} B_{c,d}^{(r)} C_{e,f}^{(r)} \quad \forall a, b, c, d, e, f),$$

where  $\times$  denotes the tensor product (equivalent to `numpy.multiply.outer()` from NumPy).

Such a decomposition yields a  $O(N^{3 \log_{nkm} R})$ -time divide-and-conquer algorithm for multiplying two  $N \times N$  matrices. [Bläser, 2013] The quantity  $3 \log_{nkm} R$  is known as the *running time exponent* and is commonly denoted  $\omega$ . The smallest  $R$  such that a  $R$ -rank decomposition of  $\mathcal{T}^{\langle n, m, k \rangle}$  exists is known as the *rank* of the tensor  $\mathcal{T}^{\langle n, m, k \rangle}$ .

---

\*Student Contributor, MIT Class of 2025, Department of Electrical Engineering and Computer Science, Cambridge, MA 02139

†Supervisor, MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA 02139

Let  $\langle n, k, m : R \rangle_{\mathbb{F}}$  be short for “a rank- $R$  decomposition of  $\mathcal{T}^{(n,k,m)}$  over field  $\mathbb{F}$ ”. [Strassen, 1969] proved that the trivial algorithm for matrix multiplication is not optimal, by finding a divide-and-conquer algorithm that can be encoded as  $\langle 2, 2, 2 : 7 \rangle_{\mathbb{Z}}$ <sup>1</sup> ( $\omega \approx 2.807$ ). The next improvement came from [Pan, 1978] with  $\langle 70, 70, 70 : 143640 \rangle_{\mathbb{Z}}$  ( $\omega \approx 2.795$ ). Every record algorithm afterward has used asymptotic inequalities to nonconstructively prove that certain running time exponents can be obtained (or approached arbitrarily closely), instead of directly finding decompositions; [Bläser, 2013] the current record achieves a running time of  $O(N^{2.37188})$ . [Duan et al., 2022]

While these asymptotic techniques allow such algorithms achieve significantly lower asymptotic running time exponents than any known decompositions of small tensors, they also render such algorithms too slow for any matrices encountered in practice; thus, there is still much demand for small, explicitly constructed decompositions. Significant progress has been made in this area in recent years, especially  $\langle 3, 3, 6 : 40 \rangle_{\mathbb{Q}}$  ( $\omega \approx 2.774$ ) [Smirnov, 2013] and  $\langle 5, 5, 5 : 97 \rangle_{\mathbb{Z}}$  ( $\omega \approx 2.842$ ) [Kauers & Moosbauer, 2022], and additional results are catalogued at <https://fmm.univ-lille.fr/>.<sup>2</sup>

We focus on  $\mathcal{T}^{(3,3,3)}$ , since it is the smallest tensor of the form  $\mathcal{T}^{(n,n,n)}$  whose rank is still unknown; the best-known upper bound is 23 ( $\omega \approx 2.854$ ), first discovered by [Laderman, 1976], which has not been beaten in 40+ years. Inspired by [Ballard et al., 2018], we enforce symmetry restrictions on decompositions to significantly reduce search space, which also motivates us to focus on tensors  $\mathcal{T}^{(n,n,n)}$ . Finally, we search over  $\mathbb{Z}_2$  to further reduce search space.

Our objective is to find (or rule out the existence of) a rank  $\leq 23$  decomposition of  $\mathcal{T}^{(3,3,3)}$  over  $\mathbb{Z}_2$ , under one of several different symmetry restrictions. We find eight decompositions of rank 23 under four different symmetry restrictions (shown in Section 5) and rule out the existence of such decompositions for many more symmetry restrictions, although we still have not found a rank 22 decomposition.

## 1.1 Notation

We will only consider tensors  $\mathcal{T}^{(n)} := \mathcal{T}^{(n,n,n)}$ .

A decomposition  $D = \{(A^{(i)}, B^{(i)}, C^{(i)})\}_i$  is a (multi)set of matrix triplets. Borrowing notation from [Kolda & Bader, 2009],  $[[D]]$  denotes  $\sum_{(A,B,C) \in D} A \times B \times C$ , which we call the “tensor valuation” of  $D$ . If  $D$  is a nested set of matrix triplets,  $[[D]]$  denotes the tensor valuation of the flattened version of  $D$ .

For a function  $f$  that transforms a matrix triplet into another matrix triplet,  $f(D)$  is short for  $\{f(d) : d \in D\}$ .

For a set of functions  $F = \{f_i\}_i, \langle f_0, f_1, \dots \rangle$  or  $\langle F \rangle$  denotes the generation of  $F$ , i.e. the set (group) of all functions that are compositions of finite (possibly empty) sequences of  $f_i$  and  $f_i^{-1}$ . For a matrix triplet  $x$ ,  $\langle F \rangle(x) = \{f(x)\}_{f \in \langle F \rangle}$  is the orbit of  $x$  under  $\langle F \rangle$ .

$\parallel$  denotes multiset sum (list concatenation but with multisets). For a (multi)set of matrix triplets  $X$ ,

<sup>1</sup>By slight abuse of notation, what we mean here is that the decomposition is over  $\mathbb{Q}$  but only uses coefficients in  $\mathbb{Z}$ .

<sup>2</sup>Interestingly, however, the current best-known running time exponent from a small tensor decomposition still seems to be  $\langle 44, 44, 44 : 36133 \rangle_{\mathbb{Q}}$  ( $\omega \approx 2.7734$ ). [Pan, 1982]

$\langle F \rangle (X) := \|\_{x \in X} \langle F \rangle (x)$ . For  $Q$  as a set of sets of matrix triplets,  $\parallel Q$  denotes  $\|\_{q \in Q} q$ , i.e. the concatenation of the elements of  $Q$ .

For a string/sequence-like object  $S$ ,  $S_{[a:b]}$  denotes the subsequence  $S_a, \dots, S_{b-1}$ , and  $S_{[:b]}$  is short for  $S_{[0:b]}$ .

Unless stated otherwise, the field  $\mathbb{F}$  we are working over will be  $\mathbb{Z}_2$ .

## 2 Symmetry and Mod 2 Constraints

We will consider three kinds of transformations on matrix triplets  $(A, B, C)$ : [De Groote, 1978]

- cycle:  $\Delta((A, B, C)) = (B, C, A)$ ;
- transpose:  $\tau((A, B, C)) = (C^\top, B^\top, A^\top)$ ;
- trace (“sandwich”):  $\phi_{X,Y,Z}((A, B, C)) = (XAY^{-1}, YBZ^{-1}, ZCX^{-1})$  for invertible  $X, Y, Z \in \mathbb{F}^{n \times n}$ ; <sup>3</sup>

It can be checked manually that for each  $f \in \{\Delta, \tau, \phi_{X,Y,Z}\}$ , if a decomposition  $D$  satisfies  $[[D]] = \mathcal{T}^{(n)}$ , then  $[[f(D)]] = \mathcal{T}^{(n)}$ . <sup>4</sup> Because of this property, we say that the tensor  $\mathcal{T}^{(n)}$  is symmetric under  $\Delta$ ,  $\tau$ , and  $\phi_{X,Y,Z}$ . By composition,  $\mathcal{T}^{(n)}$  is symmetric under any function in the group  $\Gamma := \langle \Delta, \tau, \phi_{X,Y,Z} \rangle$ .

### 2.1 Motivation for symmetry restrictions

Because  $\mathcal{T}^{(n)}$  is symmetric under each  $f \in \Gamma$ , we might guess that there exists some decomposition  $D$  of  $\mathcal{T}^{(n)}$  that happens to satisfy  $D = f(D)$  for some  $f$ ; if this is true, we say that the *decomposition*  $D$  is symmetric under  $f$ . Such a decomposition would be much easier to find than an arbitrary decomposition of  $\mathcal{T}^{(n)}$ , since there are usually far fewer symmetric decompositions than arbitrary decompositions, up to any given rank.

Furthermore, most explicitly constructed low-rank decompositions of  $\mathcal{T}^{(n)}$  happen to have some kind of symmetry. [Strassen, 1969] corresponds to  $\mathcal{T}^{(2)} = \left\langle \Delta, \phi_{F=\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, F, F} \right\rangle (\{(I_2, I_2, I_2), ([\begin{smallmatrix} 0 & 0 \\ 1 & 1 \end{smallmatrix}], [\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}], [\begin{smallmatrix} 0 & 1 \\ 0 & -1 \end{smallmatrix}])\})$  over  $\mathbb{Z}$ , and several rank-23 decompositions of  $\mathcal{T}^{(3)}$ , including the first to be discovered [Laderman, 1976], have nontrivial symmetry. [Ballard et al., 2018]

### 2.2 Motivation for working over mod 2

The most sought-after low-rank decompositions of  $\mathcal{T}^{(n)}$  are those over  $\mathbb{Z}$ , as the algorithms they yield have lower constant factors and better numerical stability, which are important in practice. Furthermore, most explicit low-rank decompositions of certain  $\mathcal{T}^{(n)}$  (and more general  $\mathcal{T}^{(n,k,m)}$ ) that have been found so far happen to be over  $\mathbb{Z}$ . [Strassen, 1969] [Laderman, 1976] [Pan, 1978] [Ballard et al., 2018]

<sup>3</sup>The  $P_\sigma$  function we used in our previous paper, for a permutation  $\sigma \in S_n$ , is equivalent to  $\phi_{S,S,S}$  where  $S$  is the corresponding permutation matrix of  $\sigma$

<sup>4</sup>The symmetries under  $\Delta$  and  $\tau$  can be shown via the identities  $\mathcal{T}_{a,b,c,d,e,f}^{(n)} = \mathcal{T}_{c,d,e,f,a,b}^{(n)}$  and  $\mathcal{T}_{a,b,c,d,e,f}^{(n)} = \mathcal{T}_{f,e,c,d,b,a}^{(n)}$ . The symmetry under  $\phi_{X,Y,Z}$  can be shown by the fact that for arbitrary matrices  $P, Q, R$ , the “dot product” of  $\mathcal{T}^{(n)}$  with  $P \times Q \times R$ ,  $\sum_{a,b,c,d,e,f} \mathcal{T}_{a,b,c,d,e,f}^{(n)} P_{a,b} Q_{c,d} R_{e,f}$ , equals  $\text{trace}(PQR)$ , which equals  $\text{trace}((XPY^{-1})(YQZ^{-1})(ZRX^{-1}))$  via algebraic properties of the trace function.

Such decompositions are extremely difficult to find. Luckily, using properties of modular arithmetic, any decomposition of  $\mathcal{T}^{(n)}$  over  $\mathbb{Z}$  must also be a decomposition over  $\mathbb{Z}_2$ , so searching mod 2 can rule out many potential decompositions over  $\mathbb{Z}$  while vastly reducing search space. Interestingly, Strassen's algorithm may have been obtained by first solving a system of equations mod 2, then extending it to  $\mathbb{Z}$ . [Landsberg, 2019]

**Disclaimer:** searching over mod 2 cannot determine anything about the existence of a non-integer decomposition, or a decomposition symmetric over some function involving non-integers, ex.  $\phi_{X=\frac{1}{2}\begin{bmatrix} -1 & \sqrt{3} \\ -\sqrt{3} & -1 \end{bmatrix}, X, X}$ .

Furthermore, if a decomposition symmetric under some functions  $f \in \Gamma$  is found over mod 2, care must be taken when extending it to  $\mathbb{Z}$ , as there may be multiple functions that are equivalent to  $f$  mod 2 but different over  $\mathbb{Z}$ , and/or the orbits of matrix triplets generated by the symmetry group of the decomposition may have different sizes in mod 2 vs. in  $\mathbb{Z}$ .

### 3 Search problem and algorithm

For some integers  $n, R$  and some set  $F \subseteq \Gamma$ , we want to find some decomposition  $D$  of rank  $\leq R$  s.t.  $[[D]] = \mathcal{T}^{(n)}$  and  $D = f(D)$  for all  $f \in F$ .

These constraints on  $D$  immediately imply  $\langle F \rangle(d) \subseteq D \ \forall d \in D$ ; thus,  $D$  must be a concatenation of orbits of individual matrix triplets under  $\langle F \rangle$ . Our search problem is then equivalent to finding a set of orbits  $Q \subseteq \Theta := \left\{ \langle F \rangle(r) : r \in (\mathbb{F}^{n \times n})^3 \right\}$  s.t.  $[[\|Q\|]] = \sum_{q \in Q} [[q]] = \mathcal{T}^{(n)}$  and  $\|Q\| = \sum_{q \in Q} |q| \leq R$ ; then  $D = \|Q\|$ . Since we are working over mod 2, WLOG we can assume  $Q$  contains distinct orbits.

#### 3.1 Nullspace compression

It can be shown that for any decomposition  $D$  and function  $f \in \{\Delta, \tau, \phi_{X,Y,Z}\}$ , applying  $f$  on  $D$  transforms  $[[D]]$  linearly: there is a matrix  $M_f$  s.t.  $\overrightarrow{[[f(D)]]} = M_f \overrightarrow{[[D]]}$  for any decomposition  $D$ , where  $\overrightarrow{T}$  denotes vectorization of a tensor  $T$ . Since linear transformations are closed under composition, this is also true for any  $f \in \Gamma$ .

Thus, for any  $T = [[q]]$  for an orbit  $q \in \Theta$ , we have  $\overrightarrow{T} = M_f \overrightarrow{T} \ \forall f \in F$   
 $\Rightarrow \overrightarrow{T} \in \text{nullspace} \left( W := \begin{bmatrix} \vdots \\ M_f - I \\ \vdots \end{bmatrix}_{f \in F} \right)$ . Performing row reduction on  $W$  yields a basis matrix  $B$  and a list

of indices  $\beta$  s.t. for any  $\overrightarrow{T} \in \text{nullspace}(W)$ ,  $\overrightarrow{T} = B \overrightarrow{v}$  for unique  $\overrightarrow{v} = \begin{bmatrix} \vdots \\ (\overrightarrow{T})_{\beta_i} \\ \vdots \end{bmatrix}_i$ ; we call  $\overrightarrow{v}$  the *compression*

of  $\overrightarrow{T} \in \text{nullspace}(W)$  and denote it as  $\overrightarrow{\overrightarrow{T}}$ . We can then replace our search condition  $\sum_{q \in Q} [[q]] = \mathcal{T}^{(n)}$  with  $\sum_{q \in Q} \overrightarrow{\overrightarrow{[[q]]}} = \overrightarrow{\overrightarrow{\mathcal{T}^{(n)}}}$  and deal with compressed tensors instead of tensors themselves.

Letting  $E_{i,j}$  denote the  $n \times n$  matrix with 1 at  $(i,j)$  and 0 everywhere else,  $\{E_{a,b} \times E_{c,d} \times E_{e,f}\}_{a,b,c,d,e,f}$

forms a basis for all 6-dimensional tensors of side length  $n$ , so  $M_f$  can be constructed by column-wise concatenating tensors  $[[f((E_{a,b}, E_{c,d}, E_{e,f}))]]$ .

### 3.2 Prefix matching

Before proceeding with the search algorithm, we choose a small number  $b$  and see if there exists a solution  $Q$  s.t.  $[[\overleftrightarrow{Q}]]$  and  $\overleftrightarrow{\mathcal{T}}^{(n)}$  match at the first  $b$  bits.

We can solve this decision problem with dynamic programming. To simplify the problem, we allow  $Q$  to have duplicate orbits.

Define  $\Lambda(r)$  denote the set of all possible  $b$ -bit prefixes of all  $Q \subseteq \Theta$  s.t.  $|Q| \leq r$ ; then we have set  $\Lambda(0) = \{\vec{0}\}$  and  $\Lambda(r) = \{v \oplus [[\overleftrightarrow{q}]]_{[:b]} : q \in \Theta, |q| \leq r, v \in \Lambda(r - |q|)\}$ . The idea is to fix one of the elements of a  $Q$  with an orbit of rank  $r$  and recurse on the rest of the orbits.

To simplify the dynamic programming, we allow  $Q$  to have duplicate orbits, so  $\Lambda(r)$  may contain unnecessary bit-vectors, but this does not affect correctness and does not significantly affect running time. The asymptotic time complexity of this procedure is  $O(2^b \cdot |\Theta| \cdot R)$ , since each  $\Lambda(r)$  must have size  $\leq 2^b$ , and at most  $|\Theta|$  many  $q$  and  $2^b$  many  $v$  for each  $q$  are considered while constructing  $\Lambda(r)$ . However, since most orbits in  $\Theta$  have relatively large rank, the actual running time is usually much lower.

After this procedure, if  $\overleftrightarrow{\mathcal{T}}^{(n)}_{[:b]} \notin \Lambda(r)$  for any  $0 \leq r \leq R$ , we know there is no solution  $Q$  for the original search problem. Due to time and memory constraints, we run this dynamic programming for each  $b = 1 \dots 25$ , terminating early if for some  $b$  we rule out the possibility of a solution to the search problem.

### 3.3 Single-orbit optimization

We collect all orbits  $q \in \Theta$  of rank  $\leq R$  and group them by their compressed tensor valuation  $[[\overleftrightarrow{q}]]$ . In each group, we choose the orbit with smallest rank, and break ties by smallest *code number*, where the code number of a matrix triplet  $(A, B, C)$  is defined as  $\$(A, B, C) := \#(A) + \#(B)2^{(n^2)} + \#(C)2^{(2n^2)}$ , with  $\#(M) := \sum_{0 \leq i < n^2} M_{\lfloor \frac{i}{n} \rfloor, i \bmod n} 2^i$ . We also remove every orbit with an all-0s compressed tensor valuation.

The remaining orbits are called *canonical*; we will only work with these orbits, since given any tensor decomposition, one can always replace each triplet with its respective representative canonical orbit, and preserve tensor valuation without increasing rank.

### 3.4 Meet-in-the-middle

We use a *meet-in-the-middle* (MITM) search strategy, which relies on the property  $[[Q||Q']] = [[Q]] + [[Q']]$ .

- Create two sets  $\check{S}, \check{S}$  s.t.  $\{\check{Q}||\check{Q} : \check{Q} \in \check{S}, \check{Q} \in \check{S}\}$  contains all  $Q$  we want to search over;<sup>5</sup>
- Store the elements of  $\check{S}$  in a dictionary  $\check{M} = \{[[\overleftrightarrow{\check{Q}}]] \rightarrow \check{Q} : \check{Q} \in \check{S}\}$  with equal keys resolved by choosing a  $\check{Q}$  with minimal rank;

---

<sup>5</sup>Unlike in our previous paper, here we only use one pair of sets instead of multiple pairs, to make the MITM simpler.

- Iterate over each  $\underline{Q} \in \underline{S}$  and query if  $v = \overrightarrow{\mathcal{T}^{(n)}} - \overrightarrow{[\underline{Q}]}$  is a key in the dictionary: if so, then  $Q = M.get(v) \parallel \underline{Q}$  is a decomposition of  $\mathcal{T}^{(n)}$ ;

### 3.4.1 Finding a meet-in-the-middle scheme

To find such  $\check{S}, \underline{S}$ , we partition the set of all  $Q$  we want to search over by what we call the *profile* of an orbit set, defined as  $\varrho(Q) = [(\#q \in Q \text{ s.t. } |q| = k)]_{0 \leq k < K}$ , where  $K = 1 + \max_{q \in \Theta} |q|$ ; a profile is essentially a histogram of orbit sizes.<sup>6</sup>

Using the fact that  $\varrho(Q \parallel Q') = \varrho(Q) + \varrho(Q')$ , where  $+$  here denotes element-wise addition, we create two sets of profiles  $\check{P}, \underline{P}$  s.t.  $\{\check{p} + \underline{p} : \check{p} \in \check{P}, \underline{p} \in \underline{P}\}$  contains all possible profiles we need to consider; then construct  $\check{S} = \{Q : \varrho(Q) \in \check{P}\}, \underline{S} = \{Q : \varrho(Q) \in \underline{P}\}$ .

Let  $H = \varrho(\Theta)$  be the histogram of orbit sizes of the set of all orbits. Since we are working over mod 2, duplicate orbits cancel each other out in tensor valuation, so WLOG we only need to consider profiles  $p$  s.t.  $p_k \leq H_k$  for each  $k$ . Let  $P$  be the set of all such  $p$  with the additional constraint  $\sum_k k p_k \leq R$  (i.e. total rank is  $\leq R$ ): this is the set of all profiles we need to consider, and it can be generated by DFS.

Given a profile  $p$ , the expression  $c(p) = \prod_{0 \leq i < K} \binom{H_k}{p_k}$  is the number of sets of orbits with profile  $p$ . In a later section we introduce methods to make a query on the dictionary  $\check{M}$  run in practically constant time. Thus, the total cost of our algorithm will be roughly proportional to  $|\check{S}| + |\underline{S}| = \sum_{p \in \check{P}} c(p) + \sum_{p \in \underline{P}} c(p)$ . We call this the “MITM cost” and denote it as  $\omega(\check{P}, \underline{P})$ . We seek to minimize it subject to the completeness constraint  $P \subseteq \{\check{p} + \underline{p} : \check{p} \in \check{P}, \underline{p} \in \underline{P}\}$  and an arbitrarily chosen memory constraint  $|\check{S}| = \sum_{p \in \check{P}} c(p) \leq m = 5 \cdot 10^8$ .

Any optimal solution can be constructed as follows: for each  $p \in P$ , choose some pair of profile  $a^{(p)}, b^{(p)} \in P$  s.t.  $a^{(p)} + b^{(p)} = p$ ; then set  $\check{P} = \{a^{(p)} : p \in P\}, \underline{P} = \{b^{(p)} : p \in P\}$ . Essentially, we specify some method of creating each  $p \in P$  using a sum-pair, then construct  $\check{P}, \underline{P}$  to contain all necessary profiles to satisfy this specification without including extra unnecessary profiles.

We use hill climbing to choose these pairs and obtain a good heuristic solution:

- Put all profiles of  $P$  in some arbitrary ordered list  $L$  and initialize  $\check{P} \leftarrow \emptyset, \underline{P} \leftarrow \emptyset$ ;
- For each  $p$  in  $L$ , find a pair  $a, b \in P$  minimizing  $\omega(\check{P} \cup \{a\}, \underline{P} \cup \{b\})$  and satisfying  $|\check{P} \cup \{a\}| \leq m$ , then add  $a$  to  $\check{P}$  and  $b$  to  $\underline{P}$ ;
- Repeat the above steps 5000 times, each time swapping a random pair of elements in  $L$  and keeping this change if it produces a solution with equal or less total cost

### 3.4.2 Fast queries

After creating the dictionary  $\check{M} := \{\overrightarrow{[\underline{Q}]} \rightarrow \check{Q} : \check{Q} \in \check{S}\}$ , we construct bitsets of all keys of  $M$  at specific chunks: specifically, we construct  $\check{\Xi}_w = \{k_{[wW:(w+1)W]} : \text{key } k \text{ in } \check{M}\}$  for each  $0 \leq w < \lceil \frac{B}{W} \rceil$ , where  $B$  is

<sup>6</sup>Note that  $\varrho(Q)_0 = 0$  since an orbit cannot have size 0.

the length of the compressed tensors,  $W = 32$  is a constant, and compressed tensors are padded with 0s as needed.

When querying whether a key  $k$  exists in  $\check{M}$ , we first query for each  $w$  whether  $k_{[wW:(w+1)W]}$  is in  $\check{\Xi}_w$  (which is done in constant time); if the answer is yes for all  $k$ , then we binary search for  $k$  among the keys of  $\check{M}$ ; otherwise we immediately know  $k$  is not in  $\check{M}$ .

In the test cases we run, only about  $\frac{1}{1000}$  or less of all queries pass these checks for all  $w$ , and we usually manage to process 10-50 million queries per second.

### 3.5 Miscellaneous Optimizations

We use multi-key quicksort to sort the keys of  $\check{M}$  by (inverse) lexicographic order (we order two bit strings  $a, b$  of equal length by their last bit, tiebreak by their next last bit, and so on).

Key-value pairs in  $\check{M}$  are stored implicitly. For each profile  $p \in \check{P}$ , we enumerate all orbit sets  $\check{Q}$  with profile  $p$  via DFS; instead of storing the entire orbit set  $\check{Q}$ , we split it into  $\kappa \cup \{q\}$ , where  $q$  is the most recent matrix triplet processed in the DFS. We collect all distinct orbit subsets  $\kappa$  encountered during the DFSs of all profiles into a list  $\mathcal{K}$ . We associate each orbit  $q \in \Theta$  with some arbitrary representative matrix triplet  $t$  s.t.  $\langle F \rangle(t) = q$ , and finally store each orbit set  $\check{Q} = \kappa \cup \{q\}$  as a number  $i(2^{3n^2}) + \$ (q)$ , where  $\mathcal{K}_i = \kappa$  and  $\$(q)$  denotes the code number of  $q$  (defined in an earlier section). As the number of elements in  $\mathcal{K}$  is usually a small fraction (usually about  $\frac{1}{1000}$ ) of the number of keys in  $\check{M}$ , storing keys this way saves a significant amount of memory. <sup>7</sup>

### 3.6 Properties of search results

Our algorithm satisfies the following lemma:

**Lemma 1.** *If there exists a decomposition  $\|Q$  of  $\mathcal{T}^{(n)}$  of rank  $\leq R$  for some  $Q \subseteq \Theta$ , then some (possibly nonequivalent) decomposition of rank  $\leq R$  must be returned by the search algorithm.*

*Proof.* Let  $Q$  be  $Q$  but with each orbit replaced with its corresponding canonical orbit. Then  $\|Q$  is a decomposition of  $\mathcal{T}^{(n)}$  with rank  $\leq \|Q\| \leq R$ , and  $Q \subseteq \Theta$ .

Let  $p = \varrho(Q)$ . By the completeness requirement of the MITM scheme, there must exist some  $\check{p} \in \check{P}, \underline{p} \in \underline{P}$  s.t.  $p = \check{p} + \underline{p}$ .  $Q$  can then be partitioned into  $\check{Q}, \underline{Q}$  s.t.  $\varrho(\check{Q}) = \check{p}$  and  $\varrho(\underline{Q}) = \underline{p}$ .

Since  $\check{M}$  must contain the compressed tensor valuations of all orbit sets with profile in  $\check{P}$ ,  $\overleftrightarrow{[[\check{Q}]}}$  must be a key contained in  $\check{M}$ ; let  $\check{Q}' = \check{M}.\text{get}(\overleftrightarrow{[[\check{Q}]})$ . Since during the construction of  $\check{M}$  we specified that equal keys are resolved by choosing a value with minimal rank,  $\|\check{Q}'\| \leq \|\check{Q}\|$ .

Since we iterate over every orbit sets with a profile in  $\underline{P}$ , at some point we will process  $\underline{Q}$ . Since  $Q = \check{Q} + \underline{Q}$ ,  $\overleftrightarrow{[[Q]]} = \overleftrightarrow{[[\check{Q}]]} + \overleftrightarrow{[[\underline{Q}]]} = \overleftrightarrow{[[\check{Q}]]} + \overleftrightarrow{[[\underline{Q}]]} = \mathcal{T}^{(n)}$ , so querying the key  $\overleftrightarrow{[[\check{Q}]]} - \overleftrightarrow{[[\underline{Q}]]}$  in  $\check{M}$  will yield the value  $\check{Q}'$ , and the algorithm will return  $Q' = \check{Q}' + \underline{Q}$ , which satisfies  $\overleftrightarrow{[[Q']]} = \mathcal{T}^{(n)}$  and  $\|Q'\| \leq \|Q\| \leq R$ .  $\square$

<sup>7</sup>This entire process is much simpler than is was in our previous paper.

Taking the contrapositive, if the search algorithm does not yield a decomposition of rank  $\leq R$ , then such a decomposition does not exist.

Since our meet-in-the-middle scheme covers all orbit sets of total rank  $\leq R$ , we can replace  $R$  in the lemma with any quantity smaller than  $R$ , and the lemma would still hold. This yields the following corollary:

**Corollary 1.** *Let  $R^*$  be the smallest possible rank for a solution  $Q$  to the search problem. If  $R^* \leq R$ , then the search algorithm is guaranteed to output a solution with rank  $R^*$ .*

Finally, we note an **important disclaimer: the search algorithm may output a solution with rank  $> R$** ; this is possible since the set  $\{\check{p} + \underline{p} : \check{p} \in \check{P}, \underline{p} \in \underline{P}\}$  may contain extraneous profiles with rank  $> R$ . This is not a problem since the algorithm is allowed to output multiple solutions, only one of which has to be optimal.

## 4 Enumerating Symmetry Subgroups

For all of our search results, we set  $n = 3$  (i.e. we searched for decompositions of  $\mathcal{T}^{(3)}$ ).

All symmetry sets  $F$  we consider will be of the form  $S \cup \{f\}$  for a small subset  $S \subseteq \mathbf{\Gamma}$  chosen beforehand and an arbitrary element  $f \in \mathbf{\Gamma}$ . For this paper we have chosen  $S = \emptyset, \{\Delta\}, \{\Delta, \top\}$ .

### 4.1 Conjugacy

Let  $G$  be the set of all functions that a given decomposition  $D$  is symmetric to. We have that  $G$  must be a group, and that conjugation by an arbitrary function that transforms matrix triplets yields  $(gGg^{-1})(gD) = gD$ , so the existence of a rank  $\leq R$  decomposition of  $\mathcal{T}^{(n)}$  symmetric over  $G$  implies the existence of rank  $\leq R$  decomposition over any conjugate of  $G$ . Thus, WLOG we only need to consider symmetry groups that are distinct up to conjugacy.

### 4.2 Optimizations

For a fixed set of functions  $S \subseteq \mathbf{\Gamma}$ , we want to enumerate all  $\langle S \cup \{f\} \rangle$  for  $f \in \mathbf{\Gamma}$  that are distinct up to conjugacy. Doing this naively by constructing  $\langle S \cup \{f\} \rangle$  for every  $f$  and testing conjugacy between pairs of such groups will be slow, as even though  $|S \cup \{f\}|$  is small,  $|\langle S \cup \{f\} \rangle|$  can be as large as  $|\mathbf{\Gamma}| = 28449792$ .<sup>8</sup> However, we can use several group-theoretic observations to take advantage of the fact that all generating sets we deal with are small, and greatly reduce memory and time usage:

- For any group  $G$  and set  $S$ ,  $G \supseteq S \Leftrightarrow G \supseteq \langle S \rangle$ ;
- For any element  $g$  and set  $S$ ,  $\langle gSg^{-1} \rangle = g \langle S \rangle g^{-1}$ ;

---

<sup>8</sup>In fact,  $\mathbf{\Gamma} = \left\langle \Delta, \phi \left[ \begin{smallmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix} \right]_{I,I} \circ \top \right\rangle$ .



These lemmas directly imply the following conjugacy tests:

- $g \langle S \rangle g^{-1} = \langle T \rangle \Leftrightarrow (gSg^{-1} \subseteq \langle T \rangle \text{ and } \langle S \rangle \supseteq g^{-1}Tg)$ ;
- $\langle S \rangle = g \langle S \rangle g^{-1} \Leftrightarrow \langle S \rangle \supseteq gSg^{-1}$  (since  $|\langle S \rangle| = |g \langle S \rangle g^{-1}|$ );

Furthermore, to detect whether  $\langle S \rangle$  and  $\langle T \rangle$  are conjugate to each other, we only need to test whether  $g \langle S \rangle g^{-1} = \langle T \rangle$  for  $g$  that are left coset representatives of  $N_{\Gamma}(\langle S \rangle)$ , the normalizer of  $\langle S \rangle$  in  $\Gamma$ .

Additionally, we do not need to iterate over every  $f$  to find all  $\langle S \cup \{f\} \rangle$  up to conjugacy, due to the following identities:

- $\langle S \cup \{f\} \rangle = \langle S \cup \{f^k\} \rangle$  for  $k$  coprime to  $\text{ord}(f)$ ;
- $\langle S \cup \{f\} \rangle = \langle S \cup \{sf\} \rangle = \langle S \cup \{fs\} \rangle$  for  $s \in \langle S \rangle$ ;
- $\eta \langle S \cup \{f\} \rangle \eta^{-1} = \langle S \cup \{\eta f \eta^{-1}\} \rangle$  for  $\eta \in N_{\Gamma}(\langle S \rangle)$ ;

Thus, after processing some  $f$ , we can transform  $f$  to  $f^k, sf, fs, \eta f \eta^{-1}$  for  $k, s, \eta$  satisfying the constraints above, then transform those elements as such and repeat until all reachable elements have been obtained, and skip these elements while enumerating subgroups.

Finally, to store the elements of  $\Gamma$  efficiently, we note that  $\tau \circ \Delta = \Delta^2 \circ \tau$ ,  $\Delta \circ \phi_{A,B,C} = \phi_{B,C,A} \circ \Delta$ , and  $\tau \circ \phi_{A,B,C} = \phi_{(A^{-1})^{\tau}, (C^{-1})^{\tau}, (B^{-1})^{\tau} \circ \tau}$ , implying  $\Gamma = \{\phi_{A,B,C} \circ \Delta^c \circ \tau^t : \text{invertible } A, B, C \in \mathbb{Z}_2^{3 \times 3}; 0 \leq c < 3; 0 \leq t < 2\}$ ; thus, each element of  $\Gamma$  can be stored with a  $O(1)$ -space canonical representation.

## 5 Solutions

For each  $S = \emptyset, \{\Delta\}, \{\Delta, \tau\}$ , the number of distinct subgroups  $\langle S \cup \{f\} \rangle$  up to conjugacy is 60, 73, and 43 respectively. For each  $F$  we tested, we initialized  $R$  to 23 and ran the dynamic programming for  $b = 1 \dots 25$ ; if no such  $b$  ruled out the possibility of a solution, we then repeatedly decremented  $R$  until the MITM cost was at most 500 billion, then ran the search algorithm for the resulting rank  $R$ .<sup>9</sup> Full search results are in Section 8.1.

Below we list all decompositions of  $\mathcal{T}^{(3)}$  that we managed to find, as well as the conjugacy relations between their symmetry subgroups. We denote  $jGj^{-1}$  as  $j \star G$ .

Several of the symmetries below happen to be conjugate to  $G_{\text{Ballard}} = \left\langle \Delta, \phi_{A=\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}, A, A} \right\rangle$ , which is the mod 2 version of  $\left\langle \Delta, \phi_{\alpha=\begin{bmatrix} 0 & 0 & -1 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \in \mathbb{Z}^{3 \times 3}, \alpha, \alpha} \right\rangle$ , under which a rank-23 decomposition of  $\mathcal{T}^{(3)}$  over  $\mathbb{Z}$  was found by [Ballard et al., 2018].

All solutions listed here except the last one have rank 23; the last solution has rank 27.

---

<sup>9</sup>We used the same computer as last time (2019 MacBook Pro, 2.3 GHz 8-Core Intel Core i9, 16 GB 2667 MHz DDR4), but a different compiler flag that offers more memory: `-Xmx30g`.



	$\left( \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \right),$ $\left( \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \right),$ $\left( \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right) \}$
--	---

$S = \{\Delta, \top\}$ :

$\langle F \rangle = \dots$	conjugacy	$D = \langle F \rangle (E)$ , where $E = \dots$
$\left\langle \Delta, \top, \phi \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, I \right\rangle$	$= \left( \phi \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \right) \star \left\langle \Delta, \phi \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, I, I \circ \top \right\rangle$	$\{ \left( \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right),$ $\left( \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right),$ $\left( \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right),$ $\left( \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \right),$ $\left( \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right) \}$ $\{ \left( \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right),$ $\left( \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right),$ $\left( \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right),$ $\left( \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \right),$ $\left( \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right) \}$
$\left\langle \Delta, \top, \phi \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\rangle$	<p>not conjugate to any other subgroup shown here;</p> <p>solution has rank 27</p>	$\{ \left( \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \right),$ $\left( \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right),$ $\left( \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \right),$ $\left( \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \right) \}$

Our source code is available at

<https://github.com/coolcomputery/Matrix-Multiplication-Tensor-Decomposition>.

## 5.1 Using flip graphs

[Kauers & Moosbauer, 2022] recently published their “flip graph” method to find low-rank mod 2 decompositions of  $\mathcal{T}^{(n,k,m)}$ , which they used to obtain  $\langle 4, 4, 4 : 47 \rangle_{\mathbb{Z}_2}$  (matching AlphaTensor [Fawzi et al., 2022]) and  $\langle 5, 5, 5 : 95 \rangle_{\mathbb{Z}_2}$ . The method repeatedly applies ‘flips’ (which change two matrix triplets at a time) and ‘reductions’ (which replace a subset of  $n$  triplets with  $n - 1$  triplets if a certain linear dependence condition is met) in a randomized manner, effectively taking a random walk in the graph where decompositions are nodes and flip/reductions are edges (the ‘flip graph’).<sup>10</sup>

<sup>10</sup>The flip graph does in fact account for symmetries of  $\mathcal{T}^{(n)}$ : it uses the equivalence class where any two decompositions that can be transformed to each other by applying some  $f \in \Gamma$  are equivalent. [Kauers & Moosbauer, 2022]

We ran the provided source code of the flip graph method on all the decompositions of  $\mathcal{T}^{(3)}$  that we found, specifying argument values `pathlength=109`, `restart=1`. The decomposition with rank 27 was only reduced to rank 23 and took about 6 hours of computation. For every other decomposition, the flip graph algorithm terminated immediately without producing any output; adding some console-printing lines in the code revealed that the algorithm was unable to do a flip or a reduction on the first step of the walk, suggesting that every such decomposition is an isolated point in the flip graph.

To confirm that this is indeed the case, we note that a flip operation transforms  $\{(A, B, C), (A, B', C')\}$  to  $\{(A, B + B', C), (A, B', C' - C)\}$  (or any index-permuted variant), so it requires both triplets to share a matrix at some index; and a reduction operation on triplets  $\{(A^{(k)}, B^{(k)}, C^{(k)})\}_k$  requires that  $\{A^{(k)}\}_k$ ,  $\{B^{(k)}\}_k$ , or  $\{C^{(k)}\}_k$  span a 1-dimensional subspace, i.e. for one of those sets to consist of matrices that are just different scalar multiples of the same underlying matrix. For each decomposition  $D$  we found that had rank 23, and for every index  $0 \leq a < 3$ , the matrices  $d_a$  for all  $d \in D$  (i.e. the  $a^{\text{th}}$  matrix of every triplet in  $D$ ) were all nonzero and mutually distinct; this immediately means no flips are possible, and since we are working over  $\mathbb{Z}_2$ , no reductions are possible either. For the last decomposition we found that had rank 27,  $|\{d_a : d \in D\}| = 21$  for each  $a$ , which explains why it could be improved with flips and reductions.

## 5.2 Takeaways from search results

We still have not found a decomposition of  $\mathcal{T}^{(3)}$  with rank 22 or lower. However, we note some interesting observations:

- Among all rank  $\leq 23$  decompositions of  $\mathcal{T}^{(3)}$  we found, there are four distinct symmetry subgroups, but only two distinct subgroups up to conjugacy, with one of them being conjugate to  $G_{\text{Ballard}}$ .
- Since none of the decompositions we found have rank  $< 23$ , by Corollary 1 this means that no such decompositions exist (over  $\mathbb{Z}_2$ ) for any of the symmetry subgroups listed above, so analyzing them further would be a dead end.
- Setting  $F = \{\text{id}\}$ ,  $R = 2$  yields no solutions, so the rank of  $\mathcal{T}^{(3)}$  is at least 3 over  $\mathbb{Z}_2$ . This is a very weak lower bound; in fact, the rank of  $\mathcal{T}^{(3)}$  is known to be at least 19 over arbitrary fields. [Bläser, 2003]
- For  $S = \{\Delta, \tau\}$ , every symmetry subgroup  $\langle S \cup \{f\} \rangle$  except for  $\langle \Delta, \tau, \text{id} \rangle$  has been proven to not have a rank  $< 23$  decomposition.
- For  $S = \{\Delta\}$ , if we restrict our attention to rank  $\leq 21$  decompositions (since 21 is the largest rank for which we would surpass Strassen's algorithm, as  $\log_3 21 \approx 2.771 < 2.807 \approx \log_2 7$ ), the only subgroups  $\langle S \cup \{f\} \rangle$  that might have a rank  $\leq 21$  decomposition are  $\langle \Delta, \text{id} \rangle$ ,  $\langle \Delta, \tau \rangle$ , and  $\left\langle \Delta, \phi_{M=\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, M, M} \right\rangle$ . The last of these subgroups seems to be the most feasible to analyze and might be worth investigating further.

## 6 Future directions

Despite our unsuccessful attempt to obtain a rank-22 decomposition of  $\mathcal{T}^{(3)}$  with our algorithm and with flip graphs, we still think that combining these two approaches is promising and that there is still much left to uncover; although [Kauers & Moosbauer, 2022] account for symmetries of  $\mathcal{T}^{(n)}$  in their equivalence classes between decompositions, they do not seem to investigate individual *decompositions* that have high amounts of symmetry.

Another strategy worth attempting could be generalizing flips and reductions to more complex transformations of decompositions, since all of our solutions except the last one with rank 27 could not be transformed at all with flips and reductions alone.

Finally, bigger tensors such as  $\mathcal{T}^{(4)}$  can be investigated for symmetric decompositions, although extra restrictions may have to be enforced on what matrix triplets are allowed, since enumerating all matrix triplets ( $\approx 2^{(3n^2)}$  of them) is only feasible on a home computer up to  $n = 3$ . One such restriction could be a sparsity constraint, such as setting a maximum allowed value of  $\mathbf{b}(A)\mathbf{b}(B)\mathbf{b}(C)$  for each triplet  $(A, B, C)$ , where  $\mathbf{b}(M)$  is the number of 1s in matrix  $M$ .

## 7 Acknowledgments

We thank Prof. Virginia Vassilevska Williams for her support throughout the past year-and-a-half, and for her suggestions to work over  $\mathbb{Z}_2$  and to test symmetry groups that superset  $\{\Delta, \tau\}$ .

## References

- Strassen, V. (1969). Gaussian elimination is not optimal. *Numerische Mathematik*, 13, pp. 354–356. Retrieved on March 24, 2023 from <https://doi.org/10.1007/BF02165411>
- Landsberg, J. M. (2019). Efficient matrix multiplication. *ILAS July 2019*. Retrieved on March 30, 2023 from <https://www.math.tamu.edu/~jml/ILAS7-19.pdf>
- Bläser, M. (2013). Fast Matrix Multiplication. *Theory of Computing Library*, Graduate Surveys 5, pp. 1-60. Retrieved on March 24, 2023 from <https://theoryofcomputing.org/articles/gs005/>
- De Groote, H. F. (1978). On varieties of optimal algorithms for the computation of bilinear mappings I. the isotropy group of a bilinear mapping. Retrieved on March 30, 2023 from [https://doi.org/10.1016/0304-3975\(78\)90038-5](https://doi.org/10.1016/0304-3975(78)90038-5)
- Ballard, G., Ikenmeyer, C., Landsberg, J. M., & Ryder, N. (2018). The Geometry of Rank Decompositions of Matrix Multiplication II:  $3 \times 3$  Matrices. Retrieved on March 24, 2023 from <https://arxiv.org/abs/1801.00843>

Duan, R., Wu, H., & Zhou, R. (2022). Faster Matrix Multiplication via Asymmetric Hashing. Retrieved on March 25, 2023 from <https://arxiv.org/abs/2210.10173>

Smirnov, A. V. (2013). The Bilinear Complexity and Practical Algorithms for Matrix Multiplication. *ISSN 09655425, Computational Mathematics and Mathematical Physics*, 2013, Vol. 53, No. 12, pp. 1781–1795. Retrieved on March 24, 2023 from <https://cs.uwaterloo.ca/~eschost/Exam/Smirnov.pdf>

Laderman, J. D. (1976). A Non-Commutative Algorithm for Multiplying  $3 \times 3$  Matrices Using 23 Multiplications. *Bulletin of the American Math Society*. 82(1), 126-128. Retrieved on March 24, 2023 from <https://www.ams.org/journals/bull/1976-82-01/S0002-9904-1976-13988-2/S0002-9904-1976-13988-2.pdf>

Pan, V. (1978). Strassen's algorithm is not Optimal Trilinear Technique of Aggregating, Uniting and Canceling for Constructing Fast Algorithms for Matrix Operations. *Proceedings of the 19th Annual Symposium on Foundations of Computer Science*. Retrieved on March 25, 2023 from <https://dl.acm.org/doi/10.1109/SFCS.1978.34>

Pan, V. (1982). Trilinear aggregating with implicit canceling for a new acceleration of matrix multiplication. Retrieved on March 30, 2023 from [https://doi.org/10.1016/0898-1221\(82\)90037-2](https://doi.org/10.1016/0898-1221(82)90037-2)

Kolda, T. & Bader, B. W. (2009). Tensor Decompositions and Applications. Retrived on March 25, 2023 from <https://doi.org/10.1137/07070111X>.

Bläser, M. (2003). On the complexity of the multiplication of matrices of small formats. *Journal of Complexity*, 19, 43-60. Retrieved on March 24, 2023 from [https://doi.org/10.1016/S0885-064X\(02\)00007-9](https://doi.org/10.1016/S0885-064X(02)00007-9).

Fawzi, A., Balog, M., Huang, A., Hubert, T., Romera, B., Barekatain, M., Novikov, A., Ruiz, F. J. R., Schrittwieser, J., Swirszcz, G., Silver, D., Hassabis, D., & Kohli, P., (2022). Discovering faster matrix multiplication algorithms with reinforcement learning. Retrieved on March 25, 2023 from <https://www.nature.com/articles/s41586-022-05172-4>

Kauers, M. & Moosbauer, J. (2022). Flip Graphs for Matrix Multiplication. Retrieved on March 25, 2023 from <https://arxiv.org/abs/2212.01175>

# Appendix

## 8.1 Full search results

$S = \emptyset$ :

$F$	min. $b \leq 25$ to rule out solutions	$R$	# sol.s	MITM cost ( $ \check{S}  +  \underline{S} $ )	MITM memory ( $ \check{S} $ )	memory time (s)	iteration time (s)
$\{\text{id}\}$	?	2	0	266865664	133432832	961.856	74.65
$\{\top\}$	?	3	0	579323768	32194	5.707	29.787
$\{\Delta\}$	?	7	0	25624652503	66673237	88.365	1027.753
$\left\{ \phi \left[ \begin{smallmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix} \right]_{I,I} \right\}$	?	2	0	55336192	2028160	7.907	12.075
$\left\{ \phi \left[ \begin{smallmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix} \right]_{I,I} \circ \top \right\}$	?	8	0	138573416984	142625869	138.131	1422.339
$\left\{ \phi \left[ \begin{smallmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix} \right]_{I,I} \circ \Delta \right\}$	?	14	0	51765372082	255497169	280.326	1627.381
$\left\{ \phi \left[ \begin{smallmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix} \right]_{I,I} \circ \Delta \circ \top \right\}$	?	7	0	2945488149	37026340	34.308	31.476
$\left\{ \phi \left[ \begin{smallmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix} \right]_{I,I} \circ \top \right\}$	1	N/A	N/A	N/A	N/A	N/A	N/A
$\left\{ \phi \left[ \begin{smallmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix} \right]_{I,I} \right\}$	?	4	0	660754173	313488281	316.1	6.474
$\left\{ \phi \left[ \begin{smallmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix} \right]_{I,I} \circ \Delta \right\}$	?	23	0	2505895612	414892095	595.849	194.488
$\left\{ \phi \left[ \begin{smallmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix} \right]_{I,I} \circ \Delta \circ \top \right\}$	?	15	0	486158755584	177016792	197.345	15304.495
$\left\{ \phi \left[ \begin{smallmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix} \right]_{I,I} \right\}$	1	N/A	N/A	N/A	N/A	N/A	N/A
$\left\{ \phi \left[ \begin{smallmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix} \right]_{I,I} \circ \Delta \right\}$	?	23	<b>2</b>	96103376	54586257	41.823	2.514
$\left\{ \phi \left[ \begin{smallmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix} \right]_{I,I} \circ \Delta \circ \top \right\}$	1	N/A	N/A	N/A	N/A	N/A	N/A
$\left\{ \phi \left[ \begin{smallmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{smallmatrix} \right]_{I,I} \right\}$	?	20	0	38123668	19061834	27.444	0.713





[illegible]

$\left\{ \phi \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\}$	25	N/A	N/A	N/A	N/A	N/A	N/A
$\left\{ \phi \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\}$	1	N/A	N/A	N/A	N/A	N/A	N/A
$\left\{ \phi \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\}$	1	N/A	N/A	N/A	N/A	N/A	N/A
$\left\{ \phi \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\}$	?	23	0	9356538	3200	0.176	0.838
$\left\{ \phi \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\}$	?	23	0	11013186	3536	0.091	0.8
$\left\{ \phi \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\}$	?	8	0	39991691337	127363705	172.737	486.564
$\left\{ \phi \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\}$	1	N/A	N/A	N/A	N/A	N/A	N/A
$\left\{ \phi \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\}$	1	N/A	N/A	N/A	N/A	N/A	N/A
$\left\{ \phi \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\}$	1	N/A	N/A	N/A	N/A	N/A	N/A
$\left\{ \phi \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \right\}$	?	8	0	38446108927	6725888	4.47	411.676
$\left\{ \phi \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \right\}$	?	20	0	37384552	18692276	48.701	2.395

$S = \{\triangle\}$ :

$F$	min. $b \leq 25$ to rule out solutions	$R$	# sol.s	MITM cost ( $ \check{S}  +  \underline{S} $ )	MITM memory ( $ \check{S} $ )	memory time (s)	iteration time (s)
$\{\triangle, \text{id}\}$	?	7	0	25624652503	66673237	88.757	1067.787
$\{\triangle, \tau\}$	?	11	0	741742362	90113122	84.064	25.17
$\left\{ \triangle, \phi \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, I, I \right\}$	?	21	0	154959229918	479827499	436.688	2851.232
$\left\{ \triangle, \phi \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, I, I \right\}^{\circ \top}$	1	N/A	N/A	N/A	N/A	N/A	N/A
$\left\{ \triangle, \phi \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, I, I \right\}^{\circ \top}$	1	N/A	N/A	N/A	N/A	N/A	N/A







[illegible]
$$S = \{\Delta, \top\}:$$

$F$	min. $b \leq 25$ to rule out solutions	$R$	# sol.s	MITM cost ( $ \check{S}  +  \underline{S} $ )	MITM memory ( $ \check{S} $ )	memory time (s)	iteration time (s)
$\{\Delta, \tau, \text{id}\}$	?	11	0	741742362	90113122	91.653	24.518
$\left\{ \Delta, \tau, \phi \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, I, I \right\}$	1	N/A	N/A	N/A	N/A	N/A	N/A
$\left\{ \Delta, \tau, \phi \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, I, I \right\}$	1	N/A	N/A	N/A	N/A	N/A	N/A
$\left\{ \Delta, \tau, \phi \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, I, I \right\}$	3	N/A	N/A	N/A	N/A	N/A	N/A
$\left\{ \Delta, \tau, \phi \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, I, I \right\}$	1	N/A	N/A	N/A	N/A	N/A	N/A
$\left\{ \Delta, \tau, \phi \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, I, I \right\}$	1	N/A	N/A	N/A	N/A	N/A	N/A
$\left\{ \Delta, \tau, \phi \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, I, I \right\}$	?	23	0	11568233	3366076	1.793	0.791
$\left\{ \Delta, \tau, \phi \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, I, I \right\}$	1	N/A	N/A	N/A	N/A	N/A	N/A
$\left\{ \Delta, \tau, \phi \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}, I, I \right\}$	1	N/A	N/A	N/A	N/A	N/A	N/A
$\left\{ \Delta, \tau, \phi \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, I, I \right\}$	3	N/A	N/A	N/A	N/A	N/A	N/A
$\left\{ \Delta, \tau, \phi \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, I \right\}$	1	N/A	N/A	N/A	N/A	N/A	N/A
$\left\{ \Delta, \tau, \phi \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, I \right\}$	1	N/A	N/A	N/A	N/A	N/A	N/A
$\left\{ \Delta, \tau, \phi \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, I \right\}$	1	N/A	N/A	N/A	N/A	N/A	N/A
$\left\{ \Delta, \tau, \phi \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}, I \right\}$	1	N/A	N/A	N/A	N/A	N/A	N/A
$\left\{ \Delta, \tau, \phi \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, I \right\}$	10	N/A	N/A	N/A	N/A	N/A	N/A
$\left\{ \Delta, \tau, \phi \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, I \right\}$	1	N/A	N/A	N/A	N/A	N/A	N/A





