THEORY OF COMPUTING, Volume 17 (2), 2021, pp. 1–32
www.theoryofcomputing.org

SPECIAL ISSUE: CCC 2019

Barriers for Fast Matrix Multiplication from Irreversibility

Matthias Christandl* Péter Vrana† Jeroen Zuiddam‡

Received July 5, 2019; Revised October 5, 2020; Published September 16, 2021

Abstract. Determining the asymptotic algebraic complexity of matrix multiplication, succinctly represented by the matrix multiplication exponent ω, is a central problem in algebraic complexity theory. The best upper bounds on ω, leading to the state-of-the-art $\omega \leq 2.37..$, have been obtained via Strassen's laser method and its generalization by Coppersmith and Winograd. Recent barrier results show limitations for these and related approaches to improve the upper bound on ω.

We introduce a new and more general barrier, providing stronger limitations than in previous work. Concretely, we introduce the notion of irreversibility of a tensor, and we prove (in some precise sense) that any approach that uses an irreversible tensor in an intermediate step (e. g., as a starting tensor in the laser method) cannot give $\omega = 2$. In quantitative terms,

A conference version of this paper appeared in the Proceedings of the 34th Computational Complexity Conference, 2019 [14].

*Supported by European Research Council (ERC Grant Agreement No. 337603 and 81876) and VILLUM FONDEN via the QMATH Centre of Excellence (Grant No. 10059).

†Supported by National Research, Development and Innovation Fund of Hungary within the Quantum Technology National Excellence Program (Project Nr. 2017-1.2.1-NKP-2017-00001) and via the research grants K124152, KH129601.

‡Supported by National Science Foundation Grant No. DMS-1638352 and indirectly supported by the National Science Foundation Grant No. CCF-1900460.

ACM Classification: F.2.1

AMS Classification: 68Q17, 68Q25

Key words and phrases: algebraic complexity, matrix multiplication, barriers, tensors, tensor rank

 DOI: 10.4086/toc.2021.v017a002

we prove that the best upper bound achievable is at least twice the irreversibility of the intermediate tensor. The quantum functionals and Strassen support functionals give (so far, the best) lower bounds on irreversibility. We provide lower bounds on the irreversibility of key intermediate tensors, including the small and big Coppersmith–Winograd tensors, that improve limitations shown in previous work. Finally, we discuss barriers on the group-theoretic approach in terms of monomial irreversibility.

1 Introduction

Determining the asymptotic algebraic complexity of matrix multiplication is a central open problem in algebraic complexity theory. Several methods for constructing fast matrix multiplication algorithms have been developed, but at a high level they typically consist of two parts: an efficient reduction of matrix multiplication to an intermediate problem (some bilinear map, i. e., some 3-tensor) and an efficient algorithm for the intermediate problem. Recent results have shown *barriers* for such constructions to yield fast matrix multiplication algorithms [4, 8, 9, 2, 3]. We give a barrier, based on a new notion called irreversibility, that is more general and in some cases stronger than the barriers from previous work.

1.1 Matrix multiplication barriers

The matrix multiplication exponent ω is defined as the infimum over all real numbers β such that any two $n \times n$ matrices can be multiplied with $\mathcal{O}(n^{\beta})$ algebraic operations, and thus ω represents the asymptotic algebraic complexity of matrix multiplication. The bounds $2 \leq \omega \leq 3$ hold trivially. Strassen published the first non-trivial upper bound $\omega \leq \log_2 7$ in 1969 [28]. In the decades that followed, through the development of several ingenious methods by various people, the upper bound was improved to the state-of-the-art bound $\omega \leq 2.37..$, and the pursuit to prove whether $\omega = 2$ or $\omega > 2$ has been ongoing [17, 27, 35, 23, 15, 16]. As mentioned before, these upper bound methods typically consist of a reduction of matrix multiplication to an intermediate problem and an efficient algorithm for the intermediate problem.

Ambainis, Filmus and Le Gall [4], for the first time, proved a barrier result for some collection of such methods. Namely, they showed that a variety of methods that go via the big Coppersmith–Winograd tensor as an intermediate problem cannot give $\omega = 2$, and in fact not even $\omega \leq 2.30...$ We call any lower bound for all upper bounds on ω that can be obtained by some method, a *barrier* for that method. In general, barriers in the sense of limitations to proof methods have a long history in computational complexity theory and recognizing barriers is a natural step towards finding proof methods that do solve the problem at hand.

Next, Alman and Williams [2, 3] extended the realm of barriers beyond the scope of the Ambainis *et al.* barrier, to a larger collection of methods. Also Blasiak, Church, Cohn, Grochow, Naslund, Sawin, and Umans [8] and Blasiak, Church, Cohn, Grochow, and Umans [9] studied barriers, namely barriers for a subset of the group-theoretic method. Both the Blasiak *et al.* and the Alman and Williams barriers rely on studying versions of asymptotic subrank of an intermediate problem.

We give a barrier that applies more generally than all previous barriers and that is in some cases stronger. Our barrier also relies on studying versions of asymptotic subrank, which together with the notion of asymptotic rank we combine into a single parameter called *irreversibility*. Our barrier simplifies and generalizes previous barriers and tightly connects the barrier literature to central notions from the framework of Strassen [29, 30, 31, 12]. Alman [1] reported very similar independent results shortly after our manuscript appeared on the arXiv, which we jointly presented at the Computational Complexity Conference 2019 in New Brunswick. For all the tensors mentioned, the barriers of Alman are identical to ours. A subtle difference between the two papers is that Alman considers asymptotic slice rank instead of asymptotic subrank.

1.2 Our barrier: informal explanation

Our barrier relies on two ideas: (i) we think of computational problems as *resources* that can be reduced to one another and (ii) such reductions satisfy a *triangle inequality* which limits the possible chains of reductions. An informal explanation of our barrier is as follows. The matrix multiplication exponent ω is the optimal *rate* at which the problem of multiplying matrices can be reduced to the problem of multiplying numbers, or in other words, the optimal rate at which the problem of multiplying numbers can be transformed into the problem of multiplying matrices,

$$\text{multiplying numbers} \xrightarrow{\omega} \text{multiplying matrices.} \tag{1.1}$$

By this we mean that for any n the problem of multiplying $n \times n$ matrices can be reduced to the problem of multiplying $n^{\omega+o(1)}$ pairs of numbers (and some additions, but they do not have an influence on the complexity). We will make these notions precise later. For now, we stick to the high-level picture. Rates of transformation naturally satisfy a triangle inequality. Therefore, upper bounds on ω can be obtained by combining the rate of transformation α_1 from the problem of multiplying numbers to some intermediate problem and the rate of transformation α_2 from the intermediate problem to the problem of multiplying matrices; this is the two-component approach alluded to earlier,

$$\text{multiplying numbers} \xrightarrow{\alpha_1} \text{intermediate problem} \xrightarrow{\alpha_2} \text{multiplying matrices.} \tag{1.2}$$

That is, $\alpha_1\alpha_2 \geq \omega$. We define the *irreversibility* of the intermediate tensor, roughly speaking, as the optimal rate of transformation from the problem of multiplying numbers to the intermediate problem *and back to the problem of multiplying numbers*. Strassen [30] (see Section 2.1) showed that the transformation rate from the matrix multiplication problem to the problem of multiplying numbers is $\frac{1}{2}$, so we can extend the chain in (1.2) to

$$\begin{aligned}\text{multiplying numbers} \xrightarrow{\alpha_1} \text{intermediate problem} \xrightarrow{\alpha_2} \text{multiplying matrices}\\ \xrightarrow{1/2} \text{multiplying numbers.}\end{aligned} \tag{1.3}$$

Using the triangle inequality again we see that $\alpha_1\alpha_2/2$ is at least the irreversibility of the intermediate problem, and hence the irreversibility of the intermediate problem provides limitations on the upper bounds $\alpha_1\alpha_2 \geq \omega$ that can be obtained from (1.2). This is formalized in Theorem 3.1.

1.3 Explicit numerical barriers

To exemplify our barrier we show that the support functionals [31] and quantum functionals [12] give (so far, the best) lower bounds on the irreversibility of the following families of intermediate problems. These intermediate problems are encoded as tensors. A tensor is a 3-dimensional array of numbers. We use $e_{i,j,k}$ to denote the tensor that is zero everywhere except for a one in coordinate (i,j,k). We will discuss tensors in more details later. The families of intermediate tensors we consider are:

- the small Coppersmith–Winograd tensors
$$\mathrm{cw}_q = \sum_{i=1}^{q} e_{0,i,i} + e_{i,0,i} + e_{i,i,0}\,,$$
- the big Coppersmith–Winograd tensors
$$\mathrm{CW}_q = e_{0,0,q+1} + e_{0,q+1,0} + e_{q+1,0,0} + \sum_{i=1}^{q} e_{0,i,i} + e_{i,0,i} + e_{i,i,0}\,,$$
- the reduced polynomial multiplication tensors
$$t_n = \sum_{\substack{i,j,k=0:\\ i+j=k}}^{n-1} e_{i,j,k}\,.$$

Our irreversibility lower bounds lead to the following explicit barriers (Section 5.1 and Section 5.2), rounded to five decimal places.

q	cw_q-barrier
2	2
3	2.02538
4	2.06244
5	2.09627
6	2.12549
7	2.15064

q	CW_q-barrier
1	2.16805
2	2.17795
3	2.19146
4	2.20551
5	2.21913
6	2.23201

n	t_n-barrier
2	2.17795
3	2.16805
4	2.15949
5	2.15237
6	2.14641
7	2.14135

In Appendix A we provide code to compute the values in these tables to arbitrary precision and for arbitrary parameters q and n. As suggested by the values in the above tables, both the cw_q-barrier and CW_q-barrier increase when q grows, and converge to 3 when q goes to infinity. The t_n-barrier, on the other hand, decreases when n grows, and converges to 2 when n goes to infinity. Families of tensors for which the barrier converges to 2 in the limit, like t_n, may be crucial in proving that $\omega = 2$.

We stress that our method can be applied to any tensor, not only the ones above. The tensors listed above play a special role in the literature and that is why we single them out. The small and big Coppersmith–Winograd tensors come from the paper [17]. They first analyze cw_5 to get an upper bound on ω, and then improve it by analyzing CW_6 and its Kronecker square. All papers from then on (Stothers, Vassilevska–Williams, Le Gall) use powers of CW_5 to get upper bounds on ω. The reduced polynomial multiplication tensors t_n give an example of a family of tensors for which the irreducibility goes to 1 when n goes to infinity and hence the barrier becomes trivial in the limit.

1.4 Comparison and other applications

Compared to [4], whose barriers apply to the laser method, our barriers are valid for a larger class of approaches (and naturally we obtain lower barriers). Compared to [3], whose barriers apply to what we call monomial degeneration, our barriers are valid for a larger class of approaches but our barriers are also higher. As a variation on our barrier we introduce a *monomial* version. Compared to [8] and [9] our monomial barriers are valid for a class of approaches that includes their simultaneous triple product property (STPP) approach, and thus we provide a uniform view on the barriers that have appeared in the literature. We have not tried to optimize the barriers that we obtain, but focus instead on introducing the barrier itself. The barrier in [1] is very similar to ours, except for using asymptotic slice rank instead of asymptotic subrank.

It will become clear to the reader during the development of our ideas that they not only apply to the problem of fast matrix multiplication, but extend to give barriers for the more general problem of constructing fast rectangular matrix multiplication algorithms (see the follow-up results on barriers for rectangular matrix multiplication in [11]) or even transformations between arbitrary powers of tensors. Such transformations may represent, for example, asymptotic SLOCC (stochastic local operations and classical communication) reductions among multipartite quantum states [6, 18, 36, 20].

We define irreversibility in Section 2. In Section 3 we discuss how irreversibility implies a barrier. In Section 4 we discuss methods to analyze irreversibility. Finally, in Section 5 we exhibit explicit irreversibility barriers.

2 Irreversibility of tensors

We begin by introducing some standard terminology and results regarding tensors and matrix multiplication. Then we discuss a useful notion called the relative exponent of two tensors and we define the irreversibility of a tensor. After that we introduce the monomial versions of these ideas. Finally, we discuss what values the irreducibility can take.

2.1 Tensors and matrix multiplication

We begin with some standard terminology and results regarding tensors and matrix multiplication. Standard references for this material are [10] and [7]. Let $\mathbb{F}$ be a field. Everything we discuss works over all fields, except when mentioned otherwise. An $n_1 \times n_2 \times n_3$ *tensor* (over $\mathbb{F}$) is a three-dimensional array

$$t = (t_{i_1,i_2,i_3})_{i_1 \in [n_1], i_2 \in [n_1], i_3 \in [n_1]}$$

of field elements $t_{i_1,i_2,i_3} \in \mathbb{F}$. We denote the set of all $n_1 \times n_2 \times n_3$ tensors by $\mathbb{F}^{n_1 \times n_2 \times n_3}$. The *support of* t is defined as the set

$$\operatorname{supp}(t) := \{(i_1,i_2,i_3) \in [n_1] \times [n_2] \times [n_3] : t_{i_1,i_2,i_3} \neq 0\}\,.$$

We use $e_{i,j,k}$ to denote the tensor that is zero everywhere except for a one in coordinate (i,j,k). For any $m_i \times n_i$ matrices A_i over $\mathbb{F}$ we define the tensor $(A_1,A_2,A_3) \cdot t \in \mathbb{F}^{m_1 \times m_2 \times m_3}$ by

$$((A_1,A_2,A_3)\cdot t)_{j_1,j_2,j_3} = \sum_{i_1,i_2,i_3} (A_1)_{j_1,i_1} (A_2)_{j_2,i_2} (A_3)_{j_3,i_3} t_{i_1,i_2,i_3}\,.$$

For $t \in \mathbb{F}^{n_1\times n_2\times n_3}$ and $s \in \mathbb{F}^{m_1\times m_2\times m_3}$ we write $t \geq s$ and say that t *restricts to* s if there are matrices A_i such that $(A_1,A_2,A_3)\cdot t = s$. We say t and s are *equivalent* if $t \geq s$ and $s \geq t$. We say t and s are *isomorphic* and write $s \cong t$ if $s = (A_1,A_2,A_3)\cdot t$ for invertible matrices A_i. We will often simplify the notation and simply write $s = t$ when s and t are isomorphic.

For $n \in \mathbb{N}$ we define the *rank-n unit tensor* $\langle n\rangle \coloneqq \sum_{i=1}^{n} e_{i,i,i} \in \mathbb{F}^{n\times n\times n}$. A *simple tensor* is a tensor that is an outer product $\sum_{i,j,k} u_i v_j w_k\, e_{i,j,k}$ of three vectors u, v and w. The *tensor rank of t* is defined as the smallest r such that t can be written as a sum of r simple tensors, that is,

$$t = \sum_{\ell=1}^{r} \sum_{i,j,k} u_i^{(\ell)} v_j^{(\ell)} w_k^{(\ell)}\, e_{i,j,k} \tag{2.1}$$

for vectors $u^{(\ell)}$, $v^{(\ell)}$ and $w^{(\ell)}$. It is crucial for us, but not hard to see, that we can phrase tensor rank in terms of the restriction preorder and diagonal tensors via $\mathrm{R}(t) = \min\{n \in \mathbb{N} : t \leq \langle n\rangle\}$. Indeed, if $t = (A_1,A_2,A_3)\cdot\langle r\rangle$, then the columns of the matrices A_1, A_2 and A_3 provide the vectors $u^{(\ell)}$, $v^{(\ell)}$ and $w^{(\ell)}$ for a tensor rank upper bound, and in the other direction such vectors $u^{(\ell)}$, $v^{(\ell)}$ and $w^{(\ell)}$ provide the matrices A_1, A_2 and A_3. What is nice about this definition is that it naturally gives rise to the following tensor parameter, which is central for our barriers: The *subrank of t* is defined as $\mathrm{Q}(t) \coloneqq \max\{n \in \mathbb{N} : \langle n\rangle \leq t\}$.

Extending the definition of Kronecker product for matrices, we define the *Kronecker product* $s \otimes t \in \mathbb{F}^{(m_1n_2)\times(m_2n_2)\times(m_3n_3)}$ for tensors $s \in \mathbb{F}^{m_1\times m_2\times m_3}$ and $t \in \mathbb{F}^{n_1\times n_2\times n_3}$ as the tensor that has coefficients given by all pairwise products of coefficients of s and t. Extending the definition of direct sum for matrices, we define the *direct sum* $s \oplus t \in \mathbb{F}^{(m_1+n_2)\times(m_2+n_2)\times(m_3+n_3)}$ as the tensor obtained by placing s and t as blocks in a block-diagonal tensor. In long expressions we will often write t^n to denote $t^{\otimes n}$. The product $\otimes$, the sum $\oplus$ and the diagonal tensors $\langle n\rangle$ behave well with respect to each other. One fact we will use frequently is that the direct sum of n copies of a tensor t is isomorphic to the Kronecker product of the diagonal tensor $\langle n\rangle$ and t.

For any tensor $t \in \mathbb{F}^{n_1\times n_2\times n_3}$ we define linear maps $t_1 : \mathbb{F}^{n_1} \to \mathbb{F}^{n_2\times n_3} : v \mapsto \sum_{i,j,k} v_i t_{i,j,k} e_{j,k}$, $t_2 : \mathbb{F}^{n_2} \to \mathbb{F}^{n_1\times n_3} : v \mapsto \sum_{i,j,k} v_j t_{i,j,k} e_{i,k}$ and $t_3 : \mathbb{F}^{n_3} \to \mathbb{F}^{n_1\times n_2} : v \mapsto \sum_{i,j,k} v_k t_{i,j,k} e_{i,j}$. These are called the *flattenings* of t. We call each function $t \mapsto \mathrm{rk}(t_i)$ a *flattening rank* of t. A basic observation and tool is that each flattening rank is multiplicative under the Kronecker product on tensors, additive under the direct sum on tensors, and monotone under the preorder $\leq$ on tensors.

The *asymptotic rank of t* is defined as

$$\underset{\sim}{\mathrm{R}}(t) \coloneqq \lim_{n\to\infty} \mathrm{R}(t^{\otimes n})^{1/n} = \inf_n \mathrm{R}(t^{\otimes n})^{1/n} \tag{2.2}$$

and the *asymptotic subrank of t* is defined as

$$\underset{\sim}{\mathrm{Q}}(t) \coloneqq \lim_{n\to\infty} \mathrm{Q}(t^{\otimes n})^{1/n} = \sup_n \mathrm{Q}(t^{\otimes n})^{1/n}. \tag{2.3}$$

We need to argue that the limits in the definitions exist. If $t \leq \langle n\rangle$ and $s \leq \langle m\rangle$, then we also have $t \otimes s \leq \langle nm\rangle$. It follows from this that tensor rank is submultiplicative under $\otimes$. Similarly, subrank is super-multiplicative under $\otimes$. Therefore, by Fekete's lemma,[1] the above limits exist and equal

[1] Fekete's lemma is as follows. Let $r_1, r_2, r_3, \ldots \in \mathbb{R}_{\geq 0}$ satisfy $r_{n+m} \leq r_n + r_m$ for all n,m. Then it holds that $\lim_{n\to\infty} r_n/n = \inf_n r_n/n$. See, e. g., [24, No. 98].

the respective infimum and supremum. We note that asymptotic rank and asymptotic subrank may equivalently be defined using *border* rank and *border* subrank, respectively, which are the approximative versions of rank and subrank (see, e. g., [7]).

For $a,b,c \in \mathbb{N}_{\geq 1}$ the *matrix multiplication tensor* $\langle a,b,c \rangle$ is defined as

$$\langle a,b,c \rangle \coloneqq \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} e_{(i,j),(j,k),(k,i)} \in \mathbb{F}^{ab \times bc \times ca}. \tag{2.4}$$

To see how $\langle n,n,n \rangle$ encodes multiplication of $n \times n$ matrices, let $E_{i,j}$ be the $n \times n$ matrix that is all-zero except for a 1 at coordinate (i,j), so that the $E_{i,j}$ form the standard basis of the vector space of $n \times n$ matrices. Note that $E_{i,j}E_{j,k} = E_{i,k}$ and that $\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} A_{(i,j)}B_{(j,k)}E_{i,k} = AB$. Thus suitable contractions of $\langle n,n,n \rangle$ with the matrices A and B produces the product AB. It is a standard result ([7, Section 4]) that the tensor rank of $\langle n,n,n \rangle$ equals, up to a constant factor, the arithmetic complexity of multiplying two $n \times n$ matrices. The argument is roughly as follows. Tensor rank upper bounds lead directly to matrix multiplication algorithms by the aforementioned contractions. On the other hand, any arithmetic matrix multiplication algorithm can be made into an arithmetic circuit with an addition layer followed by a bilinear multiplication layer followed by an addition layer, in such a way that the number of bilinear multiplication nodes is at most twice the number of multiplications in the arithmetic algorithm. The blow-up by a factor of two is because we only allow *bilinear multiplications* in our arithmetic circuit. This representation leads directly to a tensor rank upper bound. As we said before, the *matrix multiplication exponent* ω is defined as the infimum over all β for which the arithmetic complexity of multiplying two $n \times n$ matrices is at most $\mathcal{O}(n^{\beta})$. Matrix multiplication tensors are multiplicative in the sense that $\langle a,b,c \rangle \otimes \langle d,e,f \rangle$ is isomorphic to $\langle ad,be,cf \rangle$. From the multiplicativity of the matrix multiplication tensors it is easy to derive that ω is characterized by the asymptotic rank of $\langle 2,2,2 \rangle$ as follows.

Proposition 2.1 (see, e. g., [7, Theorem 5.2])**.** $\omega = \log_2 \underset{\sim}{\mathrm{R}}(\langle 2,2,2 \rangle)$.

The difficulty of determining the asymptotic rank of $\langle 2,2,2 \rangle$ is to be contrasted with the situation for the asymptotic subrank; to put it in Strassen's words: *Unlike the cynic, who according to Oscar Wilde knows the price of everything and the value of nothing, we can determine the asymptotic value of* $\langle h,h,h \rangle$ *precisely*.

Proposition 2.2 ([30])**.** $2 = \log_2 \underset{\sim}{\mathrm{Q}}(\langle 2,2,2 \rangle)$. *More precisely, for any* $h \in \mathbb{N}$ *it holds that*

$$\underset{\sim}{\mathrm{Q}}(\langle h,h,h \rangle) = h^2. \tag{2.5}$$

For completeness of the paper we give a proof of Proposition 2.2. The proof we provide here is based on Salem–Spencer sets and is slightly different from the proof in [30]. However, the two proofs become very similar when the construction of Salem–Spencer sets is unrolled.

Proof. Since each flattening rank of the diagonal tensor I_n equals n and since the flattening rank of $\langle h,h,h \rangle$ equals h^2 we find by multiplicativity and monotonicity of the flattening rank that $\underset{\sim}{\mathrm{Q}}(\langle h,h,h \rangle) \leq h^2$. We must prove the matching lower bound. Let p be a large prime. There is a subset $A \subseteq \mathbb{Z}/p\mathbb{Z}$ (a

Salem–Spencer set in $\mathbb{Z}/p\mathbb{Z}$) that is free of nontrivial three-term arithmetic progressions and that has size $|A| = p^{1-o(1)}$ [25, 5]. The support of the tensor $\langle p, p, p\rangle$ is the set of triples

$$E = \{((i, j), (j,k), (k,i)) : i, j,k \in [p]\}\,.$$

We think of E as a 3-partite 3-uniform hypergraph $E \subseteq V_1 \times V_2 \times V_3$ with vertex sets $V_\ell = [p] \times [p]$. Consider the subhypergraph $F \subseteq E$ induced by the vertex sets

$$\begin{aligned} W_1 &= \{(x,x+a) \in V_1 : x \in [p], a \in A\}\,.\\ W_2 &= \{(x,x+a) \in V_2 : x \in [p], a \in A\}\,.\\ W_3 &= \{(x+2a,x) \in V_3 : x \in [p], a \in A\}\,. \end{aligned}$$

That is, $F = E \cap (W_1 \times W_2 \times W_3)$. Then

$$F = \{((i, j), (j,k), (k,i)) : (i, j,k) = (x,x+a,x+2a) \text{ for some } x \in [p] \text{ and } a \in A\}\,.$$

Indeed, a priori, $((i, j), (j,k), (k,i)) \in F$ if and only if

$$(i, j,k) = (x,x+a,x+a+b = x+2c)$$

for some $x \in [p]$ and $a,b,c \in A$. Then $a,(a+b)/2,b$ is a three-term arithmetic progression in A, and thus $a = b$ and the claim follows. We see that the set F has $p\,|A| = p^{2-o(1)}$ elements. We also see that for any triple $((i, j), (j,k), (k,i))$ in F, the first vertex (i, j) determines the second vertex (j,k) and the third vertex (k,i), and similarly the second vertex determines the first and third vertex and the third vertex determines the first and second vertex. In other words, the edges in F are disjoint in all coordinates, that is, they form a matching. Restrict the tensor $\langle p, p, p\rangle \in \mathbb{F}^{V_1\times V_2\times V_3}$ to $\mathbb{F}^{W_1\times W_2\times W_3}$ by restricting the coordinates in each direction from V_ℓ to W_ℓ. The support of the resulting tensor is the matching F. After a suitable permutation of each W_ℓ, this tensor becomes a diagonal tensor of size $|F| = p^{2-o(1)}$. We conclude that for every prime p we have $\mathrm{Q}(\langle p, p, p\rangle) \geq p^{2-o(1)}$. By a simple tensor product argument this implies the claim for all $h \in \mathbb{N}$. Namely, for every $h \in \mathbb{N}$ we take a prime p such that h^n is not much bigger than an integer power of p, that is, $h^n \geq p^m \geq h^n/p$. Then $\langle h^n,h^n,h^n\rangle \geq \langle p^m, p^m, p^m\rangle$ and we find that $\mathrm{Q}(\langle h^n,h^n,h^n\rangle) \geq h^{2n-o(n)}$. □

Remark 2.3. We remarked before that Strassen's proof of Proposition 2.2 in [30] takes a slightly different route. Namely, it makes use of an approximative relaxation of subrank called *border subrank*, which is denoted by $\underline{\mathrm{Q}}$. He shows that $\underline{\mathrm{Q}}(\langle m,m,m\rangle) \geq \lceil \frac{3}{4}m^2\rceil$. (This lower bound, and all Strassen's lower bounds for $\underline{\mathrm{Q}}(\langle a,b,c\rangle)$, are in fact tight [21].) Strassen then relates the border subrank to the asymptotic subrank via a simple polynomial interpolation argument to get that $\underset{\sim}{\mathrm{Q}}(\langle m,m,m\rangle) = m^2$.

2.2 Relative exponent

For a clean exposition of our barrier we will use the notion of relative exponent, which we will define in this section. This notion is inspired by the notion of rate from information theory and alternatively can be seen as a versatile version of the notion of the asymptotic preorder for tensors of Strassen.[2] In the context of tensors, the relative exponent previously appeared in [38],[37] and [13].

To avoid technicalities, we will from now on, without further mentioning, only consider tensors that are not of tensor rank one or zero. This assumption is explicitly used in the proof of Theorem 2.5(i) and its applications.

Definition 2.4. For two tensors $t \in \mathbb{F}^{n_1 \times n_2 \times n_3}$ and $s \in \mathbb{F}^{m_1 \times m_2 \times m_3}$ we define the *relative exponent from t to s* as

$$\omega(t,s) := \lim_{n\to\infty} \tfrac{1}{n} \min\{m \in \mathbb{N}_{\geq 1} : t^{\otimes m} \geq s^{\otimes n}\} \tag{2.6}$$

$$= \sup_n \tfrac{1}{n} \min\{m \in \mathbb{N}_{\geq 1} : t^{\otimes m} \geq s^{\otimes n}\}. \tag{2.7}$$

The limit is a supremum by Fekete's lemma. Let us briefly relate the relative exponent to the basic notions and results stated earlier. The reader verifies directly that the identities

$$\omega(\langle 2\rangle, t) = \log_2 \underset{\sim}{\mathrm{R}}(t) \tag{2.8}$$

$$\omega(t, \langle 2\rangle) = 1/(\log_2 \underset{\sim}{\mathrm{Q}}(t)) \tag{2.9}$$

hold by approximating integers by powers of 2, or see [13, Proposition 1.1.16] for a proof. From the fact that $\omega := \log_2 \underset{\sim}{\mathrm{R}}(\langle 2,2,2\rangle)$ (Proposition 2.1) it follows that

$$\omega(\langle 2\rangle, \langle 2,2,2\rangle) = \omega\,. \tag{2.10}$$

We know from (2.5) that $\underset{\sim}{\mathrm{Q}}(\langle 2,2,2\rangle) = 4$ and so

$$\omega(\langle 2,2,2\rangle, \langle 2\rangle) = \tfrac{1}{2}\,. \tag{2.11}$$

The relative exponent has the following two basic properties.

Proposition 2.5. *Let s, t and u be tensors.*

(i) $\omega(t,t) = 1$.

(ii) $\omega(s,t)\,\omega(t,u) \geq \omega(s,u)$ (triangle inequality).

Proof. (i) Clearly $\omega(t,t) \leq 1$ by reflexivity of the restriction preorder $\geq$. Since t is not of rank 1, we can flatten t into a matrix in one of the three directions, so that the matrix rank is some number $r > 1$. It follows from multiplicativity of matrix rank, that if $t^{\otimes m} \geq t^{\otimes n}$, then $r^m \geq r^n$ and so $m \geq n$. This implies the claim. (ii) follows from the transitivity of the restriction preorder $\geq$. □

[2] The asymptotic preorder $\gtrsim$ is defined as follows: $s \gtrsim t$ if $s^{\otimes n+o(n)} \geq t^{\otimes n}$ when $n \to \infty$.

2.3 Irreversibility

Our barrier framework relies crucially on the irreversibility of a tensor, a new notion that we define now.

Definition 2.6. We define the *irreversibility* of a tensor t as the product of the relative exponent from $\langle 2\rangle$ to t and the relative exponent from t to $\langle 2\rangle$, i. e.,

$$\mathbf{i}(t) \coloneqq \omega(\langle 2\rangle, t)\,\omega(t, \langle 2\rangle)\,. \tag{2.12}$$

Thus $\mathbf{i}(t)$ measures the extent to which the asymptotic conversion from $\langle 2\rangle$ to t is irreversible, explaining the name. Equivalently, the irreversibility is the ratio of the logarithms of the asymptotic rank and the asymptotic subrank, i. e.,

$$\mathbf{i}(t) = \frac{\log_2 \underset{\sim}{\mathrm{R}}(t)}{\log_2 \underset{\sim}{\mathrm{Q}}(t)}\,. \tag{2.13}$$

Proposition 2.7. *For any tensor t it holds that*

$$\mathbf{i}(t) \geq 1\,. \tag{2.14}$$

Proof. From the basic properties of the relative exponent (Proposition 2.5) it follows directly that the inequality $\mathbf{i}(t) = \omega(\langle 2\rangle, t)\,\omega(t, \langle 2\rangle) \geq \omega(\langle 2\rangle, \langle 2\rangle) = 1$ holds. □

Definition 2.8. We call a tensor t *reversible* if $\mathbf{i}(t) = 1$ and *irreversible* otherwise (in which case it holds that $\mathbf{i}(t) > 1$ by Proposition 2.7).

For example, for any $n \in \mathbb{N}$ the diagonal tensor $\langle n\rangle = \sum_{i=1}^{n} e_{i,i,i}$ is reversible. In fact, every reversible tensor t, *that we know of*, is equivalent to $\langle n\rangle$ for some n, in the sense that $\langle n\rangle \leq t \leq \langle n\rangle$. In other words, we do not know whether every reversible tensor t is equivalent to $\langle n\rangle$ for some n.

This question is closely related to whether the matrix multiplication exponent ω equals 2. For the matrix multiplication tensor $\langle 2,2,2\rangle$ we have that $2\,\mathbf{i}(\langle 2,2,2\rangle) = \omega$ (using (2.11)). Thus $\omega = 2$ if and only if $\langle 2,2,2\rangle$ is reversible. (In fact, for any $n \in \mathbb{N}$ it holds that $\omega = 2$ if and only if $\langle n,n,n\rangle$ is reversible.) As we will see in Section 3, this is ultimately the source of our barrier.

Irreversible tensors do in fact exist. For example, the tensor $W = e_{0,0,1} + e_{0,1,0} + e_{1,0,0}$ is irreversible. Namely, it is known that $\log_2 \underset{\sim}{\mathrm{R}}(W) = 1$ and that $\log_2 \underset{\sim}{\mathrm{Q}}(W) = h(1/3) = 0.918..$ [31, Theorem 6.7], so $\mathbf{i}(W) = 1.088.. > 1$. In Section 5 we will compute lower bounds on the irreversibility of the small and big Coppersmith–Winograd tensors (which play a crucial role in the best upper bounds on ω).

2.4 Monomial relative exponent and monomial irreversibility

The following restricted version of relative exponent and irreversibility will be relevant. A *generalized subpermutation matrix* is a matrix for which every row and every column has at most one nonzero coefficient. For two tensors $t \in \mathbb{F}^{n_1 \times n_2 \times n_3}$ and $s \in \mathbb{F}^{m_1 \times m_2 \times m_3}$ we write $t \geq_{\mathrm{M}} s$ and say t *monomially restricts* to s if there are linear maps $A_i : \mathbb{F}^{n_i} \to \mathbb{F}^{m_i}$, the corresponding matrices of which are generalized subpermutation matrices in the standard basis, such that $(A_1, A_2, A_3) \cdot t = s$ [29, Section 6]. Thus s is a monomial restriction of t if we can obtain s from t by rescaling (also to zero) and permuting the slices of t. Monomial restriction plays a central role in the existing upper bounds on ω.

Replacing the preorder $\geq$ by $\geq_{\mathrm{M}}$ in Section 2 gives the notions of monomial subrank Q_{M}, monomial asymptotic subrank $\underset{\sim}{\mathrm{Q}}_{\mathrm{M}}$ and monomial relative exponent ω_{M}. (For simplicity we will use monomial restriction here, but our results will also hold with $\geq_{\mathrm{M}}$ replaced by monomial degeneration $\trianglerighteq_{\mathrm{M}}$ defined in [29, Section 6]. Monomial degeneration is the approximative version of monomial restriction. It is reflexive and transitive like monomial restriction.) Note that the notions Q_{M} and $\underset{\sim}{\mathrm{Q}}_{\mathrm{M}}$ only depend on the support of the tensor, and not on the particular values of the nonzero coefficients. We define the monomial irreversibility $\mathbf{i}_{\mathrm{M}}(t)$ of t as the product of the (normal) relative exponent from $\langle 2\rangle$ to t and the monomial relative exponent from t to $\langle 2\rangle$,

$$\mathbf{i}_{\mathrm{M}}(t) \coloneqq \omega(\langle 2\rangle, t)\, \omega_{\mathrm{M}}(t, \langle 2\rangle)\,. \tag{2.15}$$

Equivalently, we have

$$\mathbf{i}_{\mathrm{M}}(t) = \frac{\log_2 \underset{\sim}{\mathrm{R}}(t)}{\log_2 \underset{\sim}{\mathrm{Q}}_{\mathrm{M}}(t)}\,. \tag{2.16}$$

We stress that the monomial irreversibility uses the monomial asymptotic subrank, but the *regular* asymptotic rank. This is in fact what is used in practice. We note that monomial (asymptotic) rank is a useless concept, since any monomial restriction of a unit tensor is again a unit tensor. In other words, using the monomial asymptotic rank is too restrictive and gives bad algorithms from the start. In particular, monomial irreversibility depends on the tensor and not only on its support.

Proposition 2.9. *Let s, t and u be tensors.*

(i) $\omega_{\mathrm{M}}(t,t) = 1$.

(ii) $\omega_{\mathrm{M}}(s,t)\, \omega_{\mathrm{M}}(t,u) \geq \omega_{\mathrm{M}}(s,u)$ (triangle inequality).

(iii) $\omega_{\mathrm{M}}(s,t) \geq \omega(s,t)$.

(iv) $\mathbf{i}_{\mathrm{M}}(t) \geq \mathbf{i}(t)$.

Proof. The first two statements follow from reflexivity and transitivity of the monomial restriction preorder $\geq_{\mathrm{M}}$, just like in the proof of Proposition 2.5. The other two statements follow from the fact that $s \geq_{\mathrm{M}} t$ implies $s \geq t$. □

Definition 2.10. We call a tensor t *monomially reversible* if $\mathbf{i}_{\mathrm{M}}(t) = 1$ and *monomially irreversible* otherwise (in which case it holds that $\mathbf{i}_{\mathrm{M}}(t) > 1$ by Proposition 2.9 (iv) and Proposition 2.7).

There exist tensors that are reversible and monomially irreversible. We give an example.

Example 2.11. For any finite group G, we define the tensor $\langle G\rangle_{\mathbb{F}} \in \mathbb{F}^{G\times G\times G}$ by

$$\langle G\rangle_{\mathbb{F}} \coloneqq \sum_{g,h\in G} e_{g,h,gh}\,.$$

The tensor $\langle G\rangle_{\mathbb{F}}$ is often referred to as the *structure tensor* of the group algebra $\mathbb{F}[G]$. The group algebra $\mathbb{F}[G]$ is the vector space $\mathbb{F}^G$ that has a bilinear multiplication operation defined on it by setting $e_g \cdot e_h = e_{gh}$ for the basis elements $\{e_g : g \in G\}$ and extending bilinearly to $\mathbb{F}[G]$. Thus $\langle G\rangle_{\mathbb{F}}$ precisely encodes this multiplication. We stress that in the notation $\langle G\rangle_{\mathbb{F}}$ the parameter G determines the support and the parameter $\mathbb{F}$ denotes over which field we consider the tensor. Even though all coefficients of $\langle G\rangle_{\mathbb{F}}$ are 0 or 1, the choice of base field $\mathbb{F}$ determines what linear maps are allowed in the restriction preorder $\geq$ and thus the notion of equivalence and the notion of isomorphism. Crucially, for the monomial restriction preorder $\geq_{\mathrm{M}}$ applied to tensors with coefficients that are only 0 or 1, the choice of base field does not make a difference.

We consider the structure tensor

$$\langle \mathbb{Z}/3\mathbb{Z}\rangle_{\mathbb{F}} = e_{0,0,0} + e_{0,1,1} + e_{1,0,1} + e_{2,0,2} + e_{0,2,2} + e_{1,1,2} + e_{1,2,0} + e_{2,1,0} + e_{2,2,1}\,. \tag{2.17}$$

First of all, $\langle \mathbb{Z}/3\mathbb{Z}\rangle_{\mathbb{C}}$ is isomorphic to the diagonal tensor $\langle 3\rangle$. In general, for any finite abelian group A, the structure tensor $\langle A\rangle_{\mathbb{C}}$ is isomorphic to the diagonal tensor $\langle |A|\rangle$, using the Fourier transform on A to diagonalize the multiplication (see, e. g., [8] or [12] for a discussion of this). Therefore, $\underset{\sim}{\mathrm{R}}(\langle \mathbb{Z}/3\mathbb{Z}\rangle_{\mathbb{C}}) = 3$ and $\underset{\sim}{\mathrm{Q}}(\langle \mathbb{Z}/3\mathbb{Z}\rangle_{\mathbb{C}}) = 3$ and hence $\langle \mathbb{Z}/3\mathbb{Z}\rangle_{\mathbb{C}}$ is reversible.

In fact, for general $\mathbb{F}$, we have $\underset{\sim}{\mathrm{R}}(\langle \mathbb{Z}/3\mathbb{Z}\rangle_{\mathbb{F}}) = 3$ because there is a monomial degeneration from the restricted polynomial multiplication tensor to $\langle \mathbb{Z}/3\mathbb{Z}\rangle_{\mathbb{F}}$, see [12, Theorem 4.16]. On the other hand, it is known that $\underset{\sim}{\mathrm{Q}}_{\mathrm{M}}(\langle \mathbb{Z}/3\mathbb{Z}\rangle_{\mathbb{F}}) = 2.755...$ This is proven in [19, 33], see also [12] for the connection to [31]. It follows for any field $\mathbb{F}$ that $\mathbf{i}_{\mathrm{M}}(\langle \mathbb{Z}/3\mathbb{Z}\rangle_{\mathbb{F}}) = 1.08..$ and that $\langle \mathbb{Z}/3\mathbb{Z}\rangle_{\mathbb{F}}$ is monomially irreversible.

Regarding matrix multiplication, Strassen's construction for (2.11) (see also our proof of Proposition 2.2) in fact shows that

$$\omega_{\mathrm{M}}(\langle 2,2,2\rangle, \langle 2\rangle) = \tfrac{1}{2}\,. \tag{2.18}$$

2.5 Upper bounds on irreversibility

We finish this section by discussing the possible values that the irreversibility can take, as a first step towards a systematic understanding of where we may find reversible tensors or almost reversible tensors. We will see that, surprisingly at first sight, the matrix multiplication exponent ω provides bounds on the irreversibility of arbitrary tensors.

First of all, if $\underset{\sim}{\mathrm{Q}}(t) \leq 1$, then $\mathbf{i}(t) = \infty$. Thus, a priori, the irreversibility of a tensor could be any number in $[1,\infty]$. Recall that we defined the flattenings of any tensor $t \in \mathbb{F}^{n_1\times n_2\times n_3}$ as the linear maps $t_1 : \mathbb{F}^{n_1} \to \mathbb{F}^{n_2\times n_3}$, $t_2 : \mathbb{F}^{n_2} \to \mathbb{F}^{n_1\times n_3}$ and $t_3 : \mathbb{F}^{n_3} \to \mathbb{F}^{n_1\times n_2}$ that are naturally associated to t. For any tensor t that has a flattening with rank (as a linear map) at most 1 it holds that $\underset{\sim}{\mathrm{Q}}(t) \leq 1$. For example, $\underset{\sim}{\mathrm{Q}}(\langle n,1,1\rangle) = 1$. If none of the flattening ranks is at most 1, then the asymptotic subrank is at least some absolute constant $c > 1$, as follows from [30, Lemma 3.7]. Thus in that case the irreversibility is finite. On the other hand, for any $t \in \mathbb{F}^{n\times n\times n}$ the upper bound $\underset{\sim}{\mathrm{R}}(t) \leq \mathrm{R}(t) \leq n^2$ holds. We conclude that, for $t \in \mathbb{F}^{n\times n\times n}$, if any flattening rank is at most 1, then $\mathbf{i}(t) = \infty$, and otherwise

$$1 \leq \mathbf{i}(t) \leq d \log_2 n$$

for some absolute constant d. This gives us an idea of the order of magnitude of the irreversibility. We will refine this upper bound in the rest of this discussion.

As the first step towards gaining more control on the possible values of the irreversibility, we note that for asymptotic rank we have the following general upper bound.

Proposition 2.12 ([30, Proposition 3.6]). *For $t \in \mathbb{F}^{n\times n\times n}$ it holds that $\underset{\sim}{\mathrm{R}}(t) \leq n^{2\omega/3}$.*

Proof. We give a sketch of the argument. For any tensor $t \in \mathbb{F}^{n\times n\times n}$ it holds that $t \leq \langle n,n,1\rangle$ and $t \leq \langle n,1,n\rangle$ and $t \leq \langle 1,n,n\rangle$. By multiplying these inequalities, it follows that $t^{\otimes 3} \leq \langle n^2,n^2,n^2\rangle$. It is not hard to see that therefore $\underset{\sim}{\mathrm{R}}(t) \leq \underset{\sim}{\mathrm{R}}(\langle n,n,n\rangle)^{2/3} = n^{2\omega/3}$. □

It is possible that $\omega = 2$, in which case it follows from Proposition 2.12 that $\underset{\sim}{\mathrm{R}}(t) \leq n^{4/3}$ for every $t \in \mathbb{F}^{n\times n\times n}$. More generally, it is possible that for every tensor $t \in \mathbb{F}^{n\times n\times n}$ we have the highly non-trivial upper bound $\underset{\sim}{\mathrm{R}}(t) \leq n$, see [10, Problem 15.5]. Strassen conjectured this to be true for a subset of all tensors called *tight* tensors [32, Conjecture 5.3].

Proposition 2.12 gives us some control over the possible values of the irreversibility. As the next step in that direction we must understand how small the asymptotic subrank can be. First, we give an example of a tensor for which the asymptotic subrank and the asymptotic rank are far apart.

Example 2.13. Let $t \in \mathbb{F}^{n\times n\times n}$ be the tensor $t = e_{1,n,1} + e_{n,1,1} + e_{2,n,2} + e_{n,2,2} + \cdots + e_{n,n,n}$. Then $\underset{\sim}{\mathrm{Q}}(t) = 2$ while $\underset{\sim}{\mathrm{R}}(t) = n$ and so we have that $\mathbf{i}(t) = \log_2(n)$.

For the rest of the discussion we require our tensors to satisfy a property called *balanced*. This property was introduced in [30, page 121]. Not all tensors are balanced, but in any tensor space $\mathbb{F}^{n\times n\times n}$ over an algebraically closed field $\mathbb{F}$, being balanced is a generic condition in the sense of algebraic geometry. In particular, almost all elements in $\mathbb{F}^{n\times n\times n}$ are balanced. We define a tensor $t \in \mathbb{F}^{n_1\times n_2\times n_3}$ to be *balanced* if the three flattenings that we defined before are full-rank as linear maps and for each $i \in [3]$ there is an element $v \in \mathbb{F}^n$ such that the element $t_i(v)$ has full rank (as a matrix). Balanced tensors are called *1-generic tensors* in [22]. An example of a non-balanced tensor is the tensor in Example 2.13. An example of a balanced tensor is the matrix multiplication tensor $\langle n,n,n\rangle$.

Proposition 2.14. *Let $t \in \mathbb{F}^{n\times n\times n}$ be balanced. Then*

$$1 \leq \mathbf{i}(t) \leq \omega\,.$$

Proof. The claim follows from the following two ingredients. The first ingredient is the general upper bound $\underset{\sim}{\mathrm{R}}(t) \leq n^{2\omega/3}$ from Proposition 2.12. The second ingredient is the lower bound $\underset{\sim}{\mathrm{Q}}(t) \geq n^{2/3}$ [30, Proposition 3.6] for balanced tensors. We give a sketch of the argument. From the balancedness assumption it follows that $\langle n,1,1\rangle \leq t$ and $\langle 1,n,1\rangle \leq n$ and $\langle 1,1,n\rangle \leq t$. By multiplying these inequalities we get $\langle n,n,n\rangle \leq t^{\otimes 3}$. Therefore, using Proposition 2.2, we have $n^2 \leq \underset{\sim}{\mathrm{Q}}(\langle n,n,n\rangle) \leq \underset{\sim}{\mathrm{Q}}(t)^3$. □

Again, it is possible that $\omega = 2$, in which case it follows from Proposition 2.14 that $1 \leq \mathbf{i}(t) \leq 2$ for balanced t. More generally, it is possible that for every tensor $t \in \mathbb{F}^{n\times n\times n}$ we have the highly non-trivial upper bound $\underset{\sim}{\mathrm{R}}(t) \leq n$. In that case we have the following bounds.

Proposition 2.15. *Let $t \in \mathbb{F}^{n\times n\times n}$ be balanced and satisfy $\underset{\sim}{\mathrm{R}}(t) \leq n$. Then*

$$1 \leq \mathbf{i}(t) \leq 1.5\,.$$

Proof. The claim follows directly from combining $\underset{\sim}{\mathrm{Q}}(t) \geq n^{2/3}$, which follows from balancedness as we saw in the proof of Proposition 2.14, and the assumption $\underset{\sim}{\mathrm{R}}(t) \leq n$. □

Example 2.16. The upper bound in Proposition 2.15 is tight. Namely, let $t = \mathrm{cw}_q$ be the small Coppersmith–Winograd tensor with parameter q. We will see from Theorem 5.1 that $\mathbf{i}(t) \to 1.5$ when q goes to infinity.

3 Irreversibility implies barriers

With the new notion of irreversibility available, we present a barrier for approaches to upper bound ω via an intermediate tensor t. As we have discussed before, all recent successful upper bounds on ω have been obtained with constructions via an intermediate tensor. The results in this section tell us which tensors not to use as intermediate tensors and what necessary quality a good intermediate tensor has.

3.1 The irreversibility barrier

For any tensor t the inequality

$$\omega(\langle 2\rangle, t)\,\omega(t, \langle 2,2,2\rangle) \geq \omega \tag{3.1}$$

holds by the triangle inequality. Any such approach to upper bound ω respects the following barrier in terms of the irreversibility $\mathbf{i}(t)$ of t.

Theorem 3.1. *For any tensor t it holds that*

$$\omega(\langle 2\rangle, t)\,\omega(t, \langle 2,2,2\rangle) \geq 2\,\mathbf{i}(t)\,. \tag{3.2}$$

Proof. By the triangle inequality (Proposition 2.5),

$$\omega(\langle 2\rangle, t)\,\omega(t, \langle 2,2,2\rangle)\,\omega(\langle 2,2,2\rangle, \langle 2\rangle) \geq \omega(\langle 2\rangle, t)\,\omega(t, \langle 2\rangle) = \mathbf{i}(t)\,. \tag{3.3}$$

Therefore, using the fact $\omega(\langle 2,2,2\rangle, \langle 2\rangle) = \frac{1}{2}$ from (2.11), we have

$$\omega(\langle 2\rangle, t)\,\omega(t, \langle 2,2,2\rangle) \geq \frac{\mathbf{i}(t)}{\omega(\langle 2,2,2\rangle, \langle 2\rangle)} = 2\,\mathbf{i}(t)\,. \tag{3.4}$$

This proves the claim. □

Theorem 3.1, in particular, implies that if $\mathbf{i}(t) > 1$, then $\omega(\langle 2\rangle, t)\,\omega(t, \langle 2,2,2\rangle) > 2$. In other words, we cannot prove $\omega = 2$ via an irreversible intermediate tensor. However, it is possible that there exists a *sequence* of irreversible intermediate tensors with irreversibility converging to 1 that can be used to prove $\omega = 2$.

To conclude, the barrier just introduced describes what quality makes a tensor a good intermediate tensor, or put differently what intermediate tensors definitely not to use when we want to prove good upper bounds on ω. Of course the barrier does not tell us explicitly which intermediate tensor to pick, but it provides a strong *heuristic* of what to look for, namely low asymptotic rank and high asymptotic subrank.

3.2 Better barriers when the method has more structure

The barrier discussed in Section 3.1 does not tell the full story, and we will now discuss the natural and more subtle continuation. Namely, not only will we lose the game when the intermediate tensor is irreversible, we lose *even more* when we use this intermediate tensor in a catalytic fashion, meaning that our algorithm is obtained from transforming a diagonal tensor to the tensor product of a diagonal tensor and a matrix multiplication tensor (via the intermediate tensor). We will explain this more, but for now it is important to know that this strategy is commonly used in the literature (with great success). However, as we will see in this section, such a catalytic approach boosts the previous barrier even more. Thus this section gives rise to a more precise heuristic which says that, while looking for intermediate tensors with small irreversibility, we may at the same time want to think about intermediate tensors that require less use of catalysis.

Catalysis in the general context of tensors is the phenomenon that for tensors s, t and u, the inequality $s \geq t$ may be false, while the inequality $s \otimes u \geq t \otimes u$ may be true. The latter inequality is called *catalytic* with the tensor u acting as a *catalyst*.

Catalysis is widely used in matrix multiplication algorithms albeit not under this explicit terminology. We will now discuss more quantitatively what we mean when we say that an approach to upper bound ω uses catalysis. Without loss of generality we may impose that the final step of any construction of a matrix multiplication algorithm is an application of the Schönhage τ-theorem. The Schönhage τ-theorem (Strassen's general version [30]) says that

$$\underset{\sim}{\mathrm{R}}\Big(\bigoplus_{i=1}^{q} \langle a_i, b_i, c_i \rangle\Big) \geq \sum_{i=1}^{q} (a_i b_i c_i)^{\omega/3} . \tag{3.5}$$

In particular, for $a_i = b_i = c_i = a$, it holds that

$$\underset{\sim}{\mathrm{R}}\Big(\langle q \rangle \otimes \langle a, a, a \rangle\Big) \geq q a^{\omega} . \tag{3.6}$$

(And this inequality is in fact tight, although we will not go into that here.) We are using here that the direct sum of q copies of a tensor t is isomorphic to the tensor product of the diagonal tensor $\langle q \rangle$ and t. Equivalently, in the language of rates, for any $\alpha, \beta \in \mathbb{N}$ it holds that

$$\omega(\langle 2 \rangle, \langle 2 \rangle^{\alpha} \langle 2,2,2 \rangle^{\beta}) \geq \alpha + \beta\omega \tag{3.7}$$

that is

$$\frac{\omega(\langle 2 \rangle, \langle 2 \rangle^{\alpha} \langle 2,2,2 \rangle^{\beta}) - \alpha}{\beta} \geq \omega . \tag{3.8}$$

(Here α corresponds to $\log_2 q$ and β corresponds to $\log_2 a$. For simplicity and concreteness we will consider only integer α and β.) Thus for any tensor t and for any $\alpha, \beta \in \mathbb{N}$ it holds that

$$\frac{\omega(\langle 2 \rangle, t)\, \omega(t, \langle 2 \rangle^{\alpha} \langle 2,2,2 \rangle^{\beta}) - \alpha}{\beta} \geq \omega . \tag{3.9}$$

Inequality (3.9) is a method for giving upper bounds on ω and we say that it is *catalytic* when $\alpha > 0$, the catalyst being the tensor $\langle 2 \rangle^{\alpha}$. In fact, upper bounds coming from the approach in the Coppersmith–Winograd paper and its follow-ups take the form of this inequality for specific t, α and β. The following barrier in terms of α, β and the irreversibility $\mathbf{i}(t)$ of t for any method of the form (3.9) says that catalysis boosts the irreversibility barrier.

Theorem 3.2. *For any tensor t and $\alpha,\beta \in \mathbb{N}$ it holds that*

$$\frac{\omega(\langle 2\rangle,t)\,\omega(t,\langle 2\rangle^{\alpha}\langle 2,2,2\rangle^{\beta})-\alpha}{\beta} \geq 2\,\mathbf{i}(t)+\frac{\alpha}{\beta}\Big(\mathbf{i}(t)-1\Big) \geq 2\,\mathbf{i}(t)\,. \tag{3.10}$$

Proof. By the triangle inequality,

$$\omega(\langle 2\rangle,t)\,\omega(t,\langle 2\rangle^{\alpha}\langle 2,2,2\rangle^{\beta})\,\omega(\langle 2\rangle^{\alpha}\langle 2,2,2\rangle^{\beta},\langle 2\rangle) \geq \omega(\langle 2\rangle,t)\,\omega(t,\langle 2\rangle) = \mathbf{i}(t)\,. \tag{3.11}$$

Therefore,

$$\omega(\langle 2\rangle,t)\,\omega(t,\langle 2\rangle^{\alpha}\langle 2,2,2\rangle^{\beta}) \geq \frac{\mathbf{i}(t)}{\omega(\langle 2\rangle^{\alpha}\langle 2,2,2\rangle^{\beta},\langle 2\rangle)}\,. \tag{3.12}$$

We can bound this by

$$\frac{\mathbf{i}(t)}{\omega(\langle 2\rangle^{\alpha}\langle 2,2,2\rangle^{\beta},\langle 2\rangle)} \geq (\alpha+2\beta)\,\mathbf{i}(t) \tag{3.13}$$

since for any tensors s,t,u the inequality $\omega(s\otimes t,u)^{-1} \geq \omega(s,u)^{-1}+\omega(t,u)^{-1}$ holds. (The inequality in (3.13) is actually known to be an equality, since the asymptotic subrank is known to be additive on any collection of tensors of the form $\{\oplus_i t^{\otimes n_i} : n_i \in \mathbb{N}\}$ for any fixed tensor t. This for example follows from the asymptotic subrank being characterized by the asymptotic spectrum of tensors that we will discuss in the next section.) Combining (3.12) and (3.13), subtracting α, dividing by β and using that $\mathbf{i}(t)-1 \geq 0$ (Proposition 2.7) gives the barrier

$$\frac{\omega(\langle 2\rangle,t)\,\omega(t,\langle 2\rangle^{\alpha}\langle 2,2,2\rangle^{\beta})-\alpha}{\beta} \geq \frac{(\alpha+2\beta)\,\mathbf{i}(t)-\alpha}{\beta} = 2\,\mathbf{i}(t)+\frac{\alpha}{\beta}(\mathbf{i}(t)-1) \geq 2\,\mathbf{i}(t)\,. \tag{3.14}$$

This proves the claim. □

As a corollary of the above theorem, we present a barrier on any approach of the following form. The Schönhage τ-theorem implies that for any $a,b,c \in \mathbb{N}_{\geq 1}$ and any tensor t it holds that

$$\frac{\omega(\langle 2\rangle,t)\,\omega(t,\langle 2\rangle^{\alpha}\langle a,b,c\rangle)-\alpha}{\frac{1}{3}\log_2(abc)} \geq \omega\,. \tag{3.15}$$

Again, we think of this inequality as a method for upper bounding ω. To prove a barrier for this method we define a cyclic symmetrization of t. For any tensor $t=(t_{ijk}) \in \mathbb{F}^{n\times n\times n}$ we define the cyclic permutations $(1,2,3)\cdot t$ and $(1,2,3)^2\cdot t$ by cyclically permuting the indices i, j and k. We define the *cyclic symmetrization* $\mathrm{cyc}(t)$ by $\mathrm{cyc}(t) \coloneqq t\otimes((1,2,3)\cdot t)\otimes((1,2,3)^2\cdot t) \in \mathbb{F}^{n^3\times n^3\times n^3}$. For example, $\mathrm{cyc}(\langle a,b,c\rangle) = \langle a,b,c\rangle\otimes\langle c,a,b\rangle\otimes\langle b,c,a\rangle = \langle abc,abc,abc\rangle$. We prove the following barrier in terms of a,b,c, α and the irreversibility of $\mathrm{cyc}(t)$.

Corollary 3.3. *For any tensor t and $\alpha \in \mathbb{N}$ and $a,b,c \in \mathbb{N}_{\geq 1}$ it holds that*

$$\frac{\omega(\langle 2\rangle,t)\,\omega(t,\langle 2\rangle^{\alpha}\langle a,b,c\rangle)-\alpha}{\frac{1}{3}\log_2(abc)} \geq 2\,\mathbf{i}(\mathrm{cyc}(t))+\frac{\alpha}{\frac{1}{3}\log_2(abc)}(\mathbf{i}(\mathrm{cyc}(t))-1) \tag{3.16}$$

$$\geq 2\,\mathbf{i}(\mathrm{cyc}(t))\,. \tag{3.17}$$

One verifies that $\mathbf{i}(t) \geq \mathbf{i}(\mathrm{cyc}(t))$. If t is cyclically symmetric, then $\mathrm{cyc}(t) = t^{\otimes 3}$ and we have the equality $\mathbf{i}(t) = \mathbf{i}(\mathrm{cyc}(t))$.

Proof of Corollary 3.3. We first prove that

$$\omega(\langle 2\rangle, t) \geq \omega(\langle 2\rangle, \mathrm{cyc}(t)^{\frac{1}{3}}).$$

Suppose that $\langle 2\rangle^m \geq t^n$. Then also $\langle 2\rangle^m \geq ((1,2,3)\cdot t)^n$ and $\langle 2\rangle^m \geq ((1,2,3)^2 \cdot t)^n$. Multiplying these inequalities gives $\langle 2\rangle^{3m} \geq \mathrm{cyc}(t)^n$. We conclude that $m/n \geq \omega(\langle 2\rangle, \mathrm{cyc}(t)^{1/3})$. By a similar argument we find that

$$\omega(t, \langle 2\rangle^{\alpha}\langle a,b,c\rangle) \geq \omega(\mathrm{cyc}(t)^{\frac{1}{3}}, \langle 2\rangle^{\alpha}\langle 2,2,2\rangle^{\frac{1}{3}\log_2(abc)}). \tag{3.18}$$

Note that we are using real powers of tensors here inside the relative exponent $\omega(\cdot,\cdot)$. This is justified by taking powers of the relevant tensors and taking a limit. Using both inequalities and then applying Theorem 3.2 gives

$$\frac{\omega(\langle 2\rangle, t)\,\omega(t, \langle 2\rangle^{\alpha}\langle a,b,c\rangle) - \alpha}{\frac{1}{3}\log_2(abc)} \geq \frac{\omega(\langle 2\rangle, \mathrm{cyc}(t))\,\omega(\mathrm{cyc}(t), \langle 2\rangle^{\alpha}\langle 2,2,2\rangle^{\frac{1}{3}\log_2(abc)}) - \alpha}{\frac{1}{3}\log_2(abc)} \tag{3.19}$$

$$\geq 2\,\mathbf{i}(\mathrm{cyc}(t)). \tag{3.20}$$

This proves the corollary. □

3.3 Better barriers through monomial irreversibility

Finally, we impose as an extra constraint that the transformation from the intermediate tensor t to the matrix multiplication tensor happens via monomial restriction (Section 2.4), that is, we consider the approach

$$\omega(\langle 2\rangle, t)\,\omega_{\mathrm{M}}(t, \langle 2,2,2\rangle) \geq \omega \tag{3.21}$$

and the more structured approaches

$$\frac{\omega(\langle 2\rangle, t)\,\omega_{\mathrm{M}}(t, \langle 2\rangle^{\alpha}\langle 2,2,2\rangle^{\beta}) - \alpha}{\beta} \geq \omega \tag{3.22}$$

and

$$\frac{\omega(\langle 2\rangle, t)\,\omega_{\mathrm{M}}(t, \langle 2\rangle^{\alpha}\langle a,b,c\rangle) - \alpha}{\frac{1}{3}\log_2(abc)} \geq \omega. \tag{3.23}$$

The proofs in the previous sections can be directly adapted to prove:

Theorem 3.4. *For any tensor t it holds that*

$$\omega(\langle 2\rangle, t)\,\omega_{\mathrm{M}}(t, \langle 2,2,2\rangle) \geq 2\,\mathbf{i}_{\mathrm{M}}(t). \tag{3.24}$$

Theorem 3.5. *For any tensor t and $\alpha, \beta \in \mathbb{N}$ it holds that*

$$\frac{\omega(\langle 2\rangle, t)\,\omega_{\mathrm{M}}(t, \langle 2\rangle^{\alpha}\langle 2,2,2\rangle^{\beta}) - \alpha}{\beta} \geq 2\,\mathbf{i}_{\mathrm{M}}(t) + \frac{\alpha}{\beta}\Big(\mathbf{i}_{\mathrm{M}}(t) - 1\Big) \geq 2\,\mathbf{i}_{\mathrm{M}}(t). \tag{3.25}$$

Corollary 3.6. *For any tensor t and $\alpha \in \mathbb{N}$ and $a,b,c \in \mathbb{N}_{\geq 1}$ it holds that*

$$\frac{\omega(\langle 2\rangle, t)\,\omega_{\mathrm{M}}(t, \langle 2\rangle^{\alpha}\langle a,b,c\rangle) - \alpha}{\frac{1}{3}\log_2(abc)} \geq 2\,\mathbf{i}_{\mathrm{M}}(\mathrm{cyc}(t)) + \frac{\alpha}{\frac{1}{3}\log_2(abc)}(\mathbf{i}_{\mathrm{M}}(\mathrm{cyc}(t)) - 1) \tag{3.26}$$

$$\geq 2\,\mathbf{i}_{\mathrm{M}}(\mathrm{cyc}(t))\,. \tag{3.27}$$

4 Methods for lower bounding irreversibility

We have seen how lower bounds on irreversibility imply barriers. By definition of irreversibility, such lower bounds come from lower bounds on asymptotic rank and upper bounds on asymptotic subrank. In this section we discuss lower bounding irreversibility from four points of view: the asymptotic spectrum of tensors, the support functionals, the quantum functionals and the asymptotic slice rank. We will use the support functionals to compute explicit barriers in Section 5; the rest of this section serves as a survey and to provide comparison.

4.1 The asymptotic spectrum of tensors

We begin with an abstract and clean approach to characterizing irreversibility, and will move to practical bounds in the following sections. Let S be a family of tensors that is closed under $\otimes$ and $\oplus$ and that contains $\langle 1\rangle$. Strassen [30] introduced the *asymptotic spectrum of tensors* $\Delta(S)$ as the set of all maps $F : S \to \mathbb{R}_{\geq 0}$ that satisfy for any tensors s and t in S that

- $s \leq t \Rightarrow F(s) \leq F(t)$,
- $F(s \otimes t) = F(s)F(t)$,
- $F(s \oplus t) = F(s) + F(t)$,
- $F(\langle 1\rangle) = 1$.

Strassen proved in [30] that for any tensor $t \in S$ it holds that $\underset{\sim}{\mathrm{Q}}(t) = \min_{F\in\Delta(S)} F(t)$ and $\underset{\sim}{\mathrm{R}}(t) = \max_{F\in\Delta(S)} F(t)$. From this we directly obtain the following concise characterization of irreversibility in terms of the asymptotic spectrum of tensors $\Delta(S)$.

Proposition 4.1. *Let t be a tensor in S. Then*

$$\mathbf{i}(t) = \frac{\max_{F\in\Delta(S)} \log F(t)}{\min_{G\in\Delta(S)} \log G(t)}\,. \tag{4.1}$$

This characterization is clean but does not lead to a practical way of computing, or even lower bounding, the irreversibility $\mathbf{i}(t)$ for a general tensor t. This is because our knowledge of $\Delta(S)$ is very limited for any general family of tensors S. Namely, a priori we only know that for every $i \in [3]$ the function $t \mapsto \mathrm{R}(t_i)$ is in $\Delta(S)$, where t_i is the flattening as described in Section 2.5. Thus we have the inequalities $\underset{\sim}{\mathrm{Q}}(t) \leq \min_i \mathrm{R}(t_i)$ and $\underset{\sim}{\mathrm{R}}(t) \geq \max_i \mathrm{R}(t_i)$. In fact, $\max_i \mathrm{R}(t_i)$ is the best lower bound on $\underset{\sim}{\mathrm{R}}(t)$ that we know of. We will keep using this lower bound on the asymptotic rank in the coming, more practical, sections. There are, however, more powerful tools than the flattening ranks $\mathrm{R}(t_i)$ to upper bound $\underset{\sim}{\mathrm{Q}}(t)$, which we will discuss now.

4.2 The support functionals

The support functionals of Strassen provide an extremely powerful method for upper bounding the asymptotic subrank of tensors (and hence lower bounding irreversibility). The definition of the support functionals is technical at first sight, but as a reward these compact objects provide us with the best upper bounds on the asymptotic subrank. Before we define the support functionals, we need to define some notions for tensors and for probability distributions. For tensors $s,t \in \mathbb{F}^{n_1\times n_2\times n_3}$ we write $s \cong t$ and say that *s and t are isomorphic* if there are invertible linear maps A_i such that $(A_1,A_2,A_3)\cdot s = t$. Recall that we denote by

$$\mathrm{supp}(s) := \{(i_1,i_2,i_3) \in [n_1]\times[n_2]\times[n_3] : s_{i_1,i_2,i_3} \neq 0\}$$

the *support of* s. For any probability distribution Q on $[n]$ let $H(Q) = \sum_{i\in[n]} Q(i)\log_2(1/Q(i))$ denote the *Shannon entropy of* Q. For any probability distribution P on $[n_1]\times[n_2]\times[n_3]$ let P_i be the *marginal distribution on* $[n_i]$. That is, the marginal distribution P_1 is defined by $P_1(i_1) := \sum_{i_2\in[n_2]}\sum_{i_3\in[n_3]} P(i_1,i_2,i_3)$ and the other marginal distributions are defined similarly. Let $\mathcal{P}(\mathrm{supp}(s))$ denote the set of all probability distributions on $\mathrm{supp}(s)$. (Thus for any element $P \in \mathcal{P}(\mathrm{supp}(s))$ we can talk about its marginal distributions P_i and their entropies $H(P_i)$.) For any tensor t and any probability vector $\theta \in \mathbb{R}^3$ Strassen defined the *support functional* ζ^θ by setting

$$\zeta^\theta(t) := 2^{\rho^\theta(t)}$$

where

$$\rho^\theta(t) := \min_{s\cong t}\ \max_{P\in\mathcal{P}(\mathrm{supp}(s))} \sum_{i=1}^{3} \theta_i H(P_i)\,.$$

(We note that we believe that the name *support functional* derives from the general concept of *support functions* in convex analysis, rather than the support of tensors.) The minimization over all s isomorphic to t appearing in the definition of $\rho^\theta(t)$ is generally not well understood, but, fortunately, for the sake of upper bounding $\zeta^\theta(t)$ it suffices to find one good s isomorphic to t. Note that the Shannon entropy H is a concave function and thus the weighted marginal entropy $\sum_{i=1}^3 \theta_i H(P_i)$ is a concave function of P. Therefore, the analysis of the maximization over P, being a convex program over an explicitly given domain, is usually straightforward.

It is not hard to see that the three flattening ranks $\mathrm{R}(t_i)$ are among the support functionals, namely when we set θ to $(1,0,0)$, $(0,1,0)$ or $(0,0,1)$. Thus the support functionals can be thought of as interpolations between the three flattening ranks. Strassen proves in [31] the fundamental property that

$$\underset{\sim}{\mathrm{Q}}(t) \leq \zeta^\theta(t)\,.$$

There are numerous examples where $\min_\theta \zeta^\theta(t) < \min_i \mathrm{R}(t_i)$, of which we will see some later. It is not known whether the equality $\underset{\sim}{\mathrm{Q}}(t) = \min_\theta \zeta^\theta(t)$ holds in general. Strassen proved that equality holds for the family of *tight* tensors [31].[3] We conclude that we have the following lower bound on the irreducibility in terms of the flattening ranks and the support functionals:

[3] A tensor t is called *tight* if for some choice of basis there are injective maps $\alpha_1,\alpha_2,\alpha_3$ such that for every $a \in \mathrm{supp}(t)$ it holds that $\alpha_1(a_1)+\alpha_2(a_2)+\alpha_3(a_3) = 0$. Tight tensors play an important role in the laser method for constructing matrix multiplication algorithms, as the laser method requires the outer structure of the intermediate tensor to be tight.

Proposition 4.2. *Let t be a tensor. Then*

$$\mathbf{i}(t) \geq \frac{\max_i \log_2 \mathrm{R}(t_i)}{\min_\theta \rho^\theta(t)}. \tag{4.2}$$

We will use the method of the support functionals to lower bound the irreversibility of some explicit tensors in Section 5.

Since we are primarily interested in using the support functionals to upper bound the asymptotic subrank, it is worth to observe that by the von Neumann minimax theorem we may for any fixed tensor t express $\min_\theta \rho^\theta(t)$ in the concise form

$$\min_\theta \rho^\theta(t) = \min_{s \cong t} \max_{P \in \mathcal{P}(\mathrm{supp}(s))} \min_{i \in [3]} H(P_i).$$

Indeed, the function $(\theta, P) \mapsto \sum_i \theta_i H(P_i)$ is convex in θ and concave in P so the minimax theorem allows us to swap the maximization over P and the minimization over θ. Moreover, $\min_\theta \sum_i \theta_i H(P_i)$ is clearly attained when θ is one of the vertices $(1,0,0)$, $(0,1,0)$, $(0,0,1)$.

To further familiarize ourselves with the definition of ζ^θ, we discuss a simple example. For this example let t be the so-called *W-tensor* $t = e_{1,2,2} + e_{2,1,2} + e_{2,2,1}$. We will upper bound $\zeta^\theta(t)$ for $\theta = (1/3, 1/3, 1/3)$. In the evaluation of $\rho^\theta(t)$ we take s to be equal to t. (This turns out to be optimal in this case, since t is tight [31].) Then the support of s is the set

$$\mathrm{supp}(s) = \{(1,2,2), (2,1,2), (2,2,1)\}.$$

Let $P \in \mathcal{P}(\mathrm{supp}(s))$ assign probability a to $(1,2,2)$, probability b to $(2,1,2)$ and probability c to $(2,2,1)$. The function $P \mapsto \sum_i \frac{1}{3} H(P_i)$ is concave and invariant under permuting the values of a, b and c. Thus, the maximum $\max_{P \in \mathcal{P}(\mathrm{supp}(s))} \sum_i \frac{1}{3} H(P_i)$ is attained by the symmetric probability distribution P that assigns $a = b = c = 1/3$ to each element in $\mathrm{supp}(s)$. Then each marginal distribution P_i assigns probability a to 1 and probability $2a$ to 2. Thus $\max_{P \in \mathcal{P}(\mathrm{supp}(s))} \sum_i \frac{1}{3} H(P_i)$ equals the Shannon entropy of P_i, which is the *binary entropy* $h(1/3) = -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3} = 0.918\ldots$. We conclude that for the W-tensor t it holds that the support functional $\zeta^{(1/3,1/3,1/3)}(t)$ is at most $2^{h(1/3)} = 1.88\ldots$. Note that $1.88\ldots$ is strictly smaller than the flattening rank $\mathrm{R}(t_i) = 2$. We remark that in fact in this case we know the asymptotic monomial subrank to be $\underset{\sim}{\mathrm{Q}}_{\mathrm{M}}(t) = \zeta^{(1/3,1/3,1/3)}(t) = 2^{h(1/3)} = 1.88\ldots$ since this t is tight. Since a discussion of tightness is not in the scope of this paper we refer to [31] for that.

5 Explicit barriers

We exhibit explicit barriers by computing lower bounds on the irreversibility of well-known intermediate tensors that play a crucial role in the best upper bounds on the matrix multiplication exponent ω. These tensors are the small and big Coppersmith–Winograd tensors. Then we discuss the reduced polynomial multiplication tensors. Finally we discuss monomial irreversibility of structure tensors of finite group algebras and relations to the group-theoretic approach.

5.1 Irreversibility of Coppersmith–Winograd tensors

We now compute lower bounds for the irreversibility of the Coppersmith–Winograd tensors. As mentioned, we will use the support functionals of Strassen [31] in our computation to upper bound the asymptotic subrank.

(Upper bounds on the asymptotic subrank of complex tensors may be obtained, not only from the Strassen support functionals, but also from the quantum functionals. For the tensors in Theorem 5.1 and Theorem 5.4, however, it is known that the quantum functionals will give the same bound as the support functionals, since these tensors are *free tensors* [12, Section 4.3].)

Theorem 5.1 (Small Coppersmith–Winograd tensors [17, Section 6])**.** *For any integer $q \geq 2$, the irreversibility of the small Coppersmith–Winograd tensor*

$$\mathrm{cw}_q \coloneqq \sum_{i=1}^{q} e_{0,i,i} + e_{i,0,i} + e_{i,i,0} \tag{5.1}$$

satisfies

$$2\,\mathbf{i}(\mathrm{cw}_q) \geq 2 \cdot \frac{\log_2(q+1)}{\log_2 3 - \frac{2}{3} + \frac{2}{3}\log_2 q}\,. \tag{5.2}$$

Proof. The rank of each flattening of cw_q equals $q+1$. Therefore, $\underset{\sim}{\mathrm{R}}(\mathrm{cw}_q) \geq q+1$. To upper bound the asymptotic subrank $\underset{\sim}{\mathrm{Q}}(\mathrm{cw}_q)$ one can upper bound the Strassen support functional with $\theta = (1/3, 1/3, 1/3)$ as in [12, Example 4.22] by

$$\rho^{\theta}(\mathrm{cw}_q) \leq \log_2 3 - \frac{2}{3} + \frac{2}{3}\log_2 q\,. \tag{5.3}$$

We find that

$$\mathbf{i}(\mathrm{cw}_q) \geq \frac{\log_2(q+1)}{\log_2 3 - \frac{2}{3} + \frac{2}{3}\log_2 q}\,. \tag{5.4}$$

This proves the theorem. □

Remark 5.2. If $q > 2$, then the right-hand side of (5.2) is at least 2.02.. See the table in Section 1 for more values. If $q = 2$, however, then the right-hand side of (5.2) equals 2. Theorem 5.1 thus does not rule out using cw_2 to prove that $\omega = 2$. Indeed, as observed in [17, Section 11]), if $\omega(\langle 2\rangle, \mathrm{cw}_2) = \log_2 3$, then $\omega = 2$.

Currently, the best upper bound we have on $\omega(\langle 2\rangle, \mathrm{cw}_q)$ is $\log_2(q+2)$. If $\omega(\langle 2\rangle, \mathrm{cw}_q) = \log_2(q+2)$, then instead of (5.2) we get the better barrier

$$2\,\mathbf{i}(\mathrm{cw}_q) \geq \frac{2\log_2(q+2)}{\log_2 3 - \frac{2}{3} + \frac{2}{3}\log_2 q}\,. \tag{5.5}$$

The right-hand side of (5.5) has a minimum value of

$$\frac{18}{5\log_2 3} = 2.27.. \tag{5.6}$$

attained at $q = 6$.

Remark 5.3. The following computation serves as a sanity check for our barrier given by Theorem 3.2. Namely we see in an example how by putting some extra assumption the barrier given by Theorem 3.2 becomes tight. Coppersmith and Winograd in [17] used cw_q as an intermediate tensor in combination with the laser method and a certain *outer structure*. Here outer structure refers to the idea of viewing the intermediate tensor as a block tensor. The outer structure is the block-support of the intermediate tensor, see also [7, Section 9]. For cw_q the outer structure is the W-tensor that we discussed in Section 4.2 which has asymptotic subrank $2^{h(1/3)}$. When we impose that we apply the laser method on cw_q with this outer structure to get an upper bound $\hat{\omega} \geq \omega$ we get the following better barrier via Theorem 3.2 or Corollary 3.3 and $\alpha = h(1/3)$ and $\beta = \frac{1}{3}\log_2(q)$:

$$\hat{\omega} \geq 2\,\mathbf{i}(\mathrm{cw}_q) + \frac{h(1/3)}{\frac{1}{3}\log_2(q)}(\mathbf{i}(\mathrm{cw}_q) - 1)\,. \tag{5.7}$$

Some values of (5.7) are as follows, using (5.4) to get a bound on $\mathbf{i}(\mathrm{cw}_q)$:

q	
2	2
3	2.04744
4	2.10545
5	2.15338
6	2.19236
7	2.22455

If in addition we assume that $\omega(\langle 2\rangle, \mathrm{cw}_q) = \log_2(q+2)$, then we obtain the barrier

$$\hat{\omega} \geq 2\,\mathbf{i}(\mathrm{cw}_q) + \frac{h(1/3)}{\frac{1}{3}\log_2(q)}(\mathbf{i}(\mathrm{cw}_q) - 1)\,. \tag{5.8}$$

Some values of (5.8) are as follows, using (5.5) to get a bound on $\mathbf{i}(\mathrm{cw}_q)$:

q	
2	3.24511
3	2.65678
4	2.50000
5	2.44072
6	2.41594
7	2.40614
8	2.40363
9	2.40492
10	2.40824
11	2.41266

with minimum value of 2.40... These barriers in fact match the upper bound

$$\omega \leq \log_q \frac{4(q+2)^3}{27} \tag{5.9}$$

that was obtained by Coppersmith and Winograd by applying the laser method in the way described above. Thus our sanity check succeeds. Other intermediate tensors with a given outer structure may be analyzed similarly.

Theorem 5.4 (Big Coppersmith–Winograd tensors [17, Section 7])**.** *For any integer $q \geq 1$ the irreversibility of the big Coppersmith–Winograd tensor*

$$\mathrm{CW}_q := e_{0,0,q+1} + e_{0,q+1,0} + e_{q+1,0,0} + \sum_{i=1}^{q} e_{0,i,i} + e_{i,0,i} + e_{i,i,0} \tag{5.10}$$

satisfies

$$2\,\mathbf{i}(\mathrm{CW}_q) \geq \begin{cases} \dfrac{2\log_2(3)}{f_1(\frac{1}{18}(\sqrt{33}-3))} = 2.16.. & \text{if } q=1 \\[2ex] \dfrac{2\log_2(4)}{f_2(\frac{1}{9})} = 2.17.. & \text{if } q=2 \\[2ex] \dfrac{2\log_2(q+2)}{f_q\left(\frac{3q-\sqrt{32+q^2}}{6(q^2-4)}\right)} & \text{if } q \geq 3 \end{cases} \tag{5.11}$$

where

$$f_q(x) := -\left(\frac{2}{3} - qx\right)\log_2\left(\frac{2}{3} - qx\right) - q \cdot 2x\log_2(2x) - \left(\frac{1}{3} - qx\right)\log_2\left(\frac{1}{3} - qx\right). \tag{5.12}$$

Proof. First, we compute the asymptotic rank of CW_q. One verifies directly that the matrix rank of each flattening of CW_q equals $q+2$, so $q+2 \leq \underset{\sim}{\mathrm{R}}(\mathrm{CW}_q)$. On the other hand, the well-known border rank upper bound $\underline{\mathrm{R}}(\mathrm{CW}_q) \leq q+2$ implies that $\underset{\sim}{\mathrm{R}}(\mathrm{CW}_q) \leq q+2$. We conclude that $\underset{\sim}{\mathrm{R}}(\mathrm{CW}_q) = q+2$.

Second, we will upper bound the asymptotic subrank $\underset{\sim}{\mathrm{Q}}(\mathrm{CW}_q)$ via the Strassen support functional $\zeta^\theta(\mathrm{CW}_q)$ with $\theta = (1/3, 1/3, 1/3)$. Recall that $\zeta^\theta(\mathrm{CW}_q) = 2^{\rho^\theta(\mathrm{CW}_q)}$ where

$$\rho^\theta(\mathrm{CW}_q) = \min_{s \cong \mathrm{CW}_q} \max_{P \in \mathcal{P}(\mathrm{supp}(s))} \sum_{i=1}^{3} \theta_i H(P_i)\,.$$

We let $s = \mathrm{CW}_q$. Then the support of CW_q is given by

$$\mathrm{supp}(\mathrm{CW}_q) = \{(0,i,i),(i,0,i),(i,i,0) : i \in [q]\} \cup \{(0,0,q+1),(0,q+1,0),(q+1,0,0)\}\,. \tag{5.13}$$

We will now evaluate $\max_{P \in \mathcal{P}(\mathrm{supp}(\mathrm{CW}_q))} \sum_{i=1}^{3} \theta_i H(P_i)$. We first observe that $\mathrm{supp}(\mathrm{CW}_q)$ is symmetric under permutation of the three coordinates and under permutation of the elements of $[q]$. The function $P \mapsto \sum_{i=1}^{3} \theta_i H(P_i)$ is concave. Thus by a simple convexity argument we may in our maximization, without loss of generality, consider only distributions P that assign some probability x to each of $(0,i,i)$, $(i,0,i)$ and $(0,i,i)$ for all $i \in [q]$, and probability $\frac{1}{3} - qx$ to each of $(0,0,q+1)$, $(0,q+1,0)$ and $(q+1,0,0)$. Under this simplification, we have that the average marginal entropy $\sum_{i=1}^{3} \theta_i H(P_i)$ equals $f_q(x)$ as defined in the theorem statement. It remains to compute the maximum of $f_q(x)$ over all $0 \leq x$ for which $3qx \leq 1$.

From a calculus computation (easy to verify using computer algebra software) it follows that the maximum of $f_q(x)$ over all $0 \leq x$ for which $3qx \leq 1$ is attained at

$$x_0 = \begin{cases} \frac{1}{18}(\sqrt{33}-3) & \text{if } q = 1 \\ \frac{1}{9} & \text{if } q = 2 \\ \dfrac{3q - \sqrt{32+q^2}}{6(q^2-4)} & \text{if } q \geq 3\,. \end{cases} \tag{5.14}$$

Then $\underset{\sim}{\mathrm{Q}}(\mathrm{CW}_q) \leq 2^{f_q(x_0)}$. Thus we have

$$2\,\mathbf{i}(\mathrm{CW}_q) \geq 2\log_2 \underset{\sim}{\mathrm{R}}(\mathrm{CW}_q)/\log_2 \underset{\sim}{\mathrm{Q}}(\mathrm{CW}_q) \geq 2\log_2(q+2)/f_q(x_0)\,,$$

which evaluates to the bound in the claim. □

Remark 5.5. The lowest value of the right-hand side of (5.11) is 2.16.. attained at $q = 1$. See the table in Section 1 for more values and see Appendix A for code to compute any values.

5.2 Irreversibility of reduced polynomial multiplication tensors

The reduced polynomial multiplication tensors t_n are an example of a natural family of tensors in which each tensor is irreversible, but where the irreversibility converges to 1 when n goes to infinity. The irreversibility of the tensors t_n can be computed directly from the computation of the asymptotic rank and asymptotic subrank of t_n in [31, Theorem 6.7], which uses the support functionals. Namely, Strassen shows that

$$\underset{\sim}{\mathrm{R}}(t_n) = n$$

and

$$\underset{\sim}{\mathrm{Q}}(t_n) = z(n)$$

where

$$z(n) = \frac{g^n - 1}{g - 1} g^{-2(n-1)/3}$$

and $g > 1$ is the unique positive real solution to the equation

$$\frac{1}{g-1} - \frac{n}{g^n - 1} = \frac{n-1}{3}\,.$$

Thus

$$\mathbf{i}(t_n) = \frac{\log_2(n)}{\log_2 z(n)}\,.$$

The limit $\lim_{n\to\infty} z(n)/n$ equals a constant, namely 0.84143... (see, e. g., [8, Equation 4.11]). It follows that

$$\lim_{n\to\infty} \mathbf{i}(t_n) = \lim_{n\to\infty} \log_2(n)/\log_2 z(n) = 1\,.$$

Thus t_n is "reversible in the limit." In Appendix A we provide code to compute the values of $\mathbf{i}(t_n)$ for any n.

5.3 Monomial irreversibility of structure tensors of finite group algebras

We now discuss irreversibility and monomial irreversibility in the context of the group-theoretic approach developed in [15]. This approach produces upper bounds on ω via intermediate tensors that are structure tensors of complex group algebras of finite groups. As defined earlier, let $\langle G\rangle_{\mathbb{F}}$ denote the structure tensor of the group algebra $\mathbb{F}[G]$ of the finite group G, in the standard basis, that is,

$$\langle G\rangle_{\mathbb{F}} \coloneqq \sum_{g,h\in G} e_{g,h,gh} \in \mathbb{F}^{G\times G\times G}. \tag{5.15}$$

The group-theoretic approach (in particular [15, Theorem 4.1]) produces an inequality of the form

$$\langle G\rangle_{\mathbb{C}} \geq_{\mathrm{M}} \langle a,b,c\rangle \tag{5.16}$$

(the base field that is used in [15] is $\mathbb{C}$, in order to make the representation theory work) which ultimately (see [15, Eq. (1)]) leads to the bound

$$\frac{\omega(\langle 2\rangle, \langle G\rangle_{\mathbb{C}})\, \omega_{\mathrm{M}}(\langle G\rangle_{\mathbb{C}}, \langle a,b,c\rangle)}{\frac{1}{3}\log_2(abc)} \geq \omega \tag{5.17}$$

where $\geq_{\mathrm{M}}$ and ω_{M} are the monomial restriction and monomial relative exponent defined in Section 2.4.

Now the monomial irreversibility barrier from Section 3.3 comes into play. Upper bounds on the monomial asymptotic subrank of $\langle G\rangle_{\mathbb{F}}$ have (using different terminology) been obtained in [8, 9, 26]. Those upper bounds imply that $\langle G\rangle_{\mathbb{F}}$ is monomially irreversible for every nontrivial finite group G. Together with our results in Section 3.3 and the fact that the tensor $\langle G\rangle_{\mathbb{F}}$ is cyclically symmetric up to a permutation of the basis of one of the tensor legs, this directly leads to nontrivial barriers for the left-hand side of (5.17) for any fixed nontrivial group G, thus putting the methods of [8, 9, 26] in a broader context. We have not tried to numerically optimize the monomial irreversibility barriers for group algebras. We gave an example of a monomial barrier for $G = \mathbb{Z}/3\mathbb{Z}$ in Example 2.11.

It should be stressed again that although these results rule out using monomial restrictions from powers of any one single group, part of the hope of the group-theoretic approach is to use a family of groups (such as the symmetric group S_n for increasing n) which is *not* just powers of a single starting group. For families of abelian groups of bounded exponent (even if they are not powers of a single group), [8] rules out STPP constructions reaching $\omega = 2$, and for certain families of nilpotent groups (again, even if they are not just powers of a single group) [9] does similarly.

The results of [8], [9] and [26], in fact, show slightly more than the aforementioned barriers for monomial restriction. While, over arbitrary fields, they only rule out monomial restrictions, over fields of bad characteristic (e. g., characteristic p when G is a p-group satisfying the relevant conditions of their theorems) they also rule out *arbitrary degenerations*. This is because they show slice rank upper bounds, the slice rank of a diagonal tensor equals its rank, and having slice rank at most r is a Zariski-closed condition (proved in [34]).

Finally, we mention that the irreversibility barrier (rather than the monomial irreversibility barrier) does not rule out obtaining $\omega = 2$ via $\langle G\rangle_{\mathbb{C}}$. Namely, $\langle G\rangle_{\mathbb{C}}$ is isomorphic to a direct sum of matrix multiplication tensors, $\langle G\rangle_{\mathbb{C}} \cong \bigoplus_i \langle d_i, d_i, d_i\rangle$ and, therefore, the irreversibility satisfies $\mathbf{i}(\langle G\rangle_{\mathbb{C}}) = (\log_2 \sum_i d_i^{\omega})/(\log_2 \sum_i d_i^2)$. Thus, if $\omega = 2$, then $\langle G\rangle_{\mathbb{C}}$ is reversible.

6 Outlook for further barriers

In Section 5 we used the support functionals of Section 4 to upper bound the asymptotic subrank and thus lower bound the irreversibility of intermediate tensors. There are two other approaches to do this that we will discuss now. These approaches are at least as powerful as the support functionals for upper bounding asymptotic subrank, and we expect that there are examples where they perform better.

6.1 The quantum functionals

For tensors over the complex numbers (i. e., when $\mathbb{F} = \mathbb{C}$) we have a much deeper understanding of the theory of upper bounds on the asymptotic subrank. For any probability vector $\theta \in \mathbb{R}^3$ the *quantum functional* F^θ introduced in [12], or rather its logarithm, is defined as

$$\log_2 F^\theta(t) = \max_{P \in \Pi(t)} \sum_{i=1}^{r} \theta_i H(P_i)$$

where $\Pi(t)$ is the *moment polytope of* t. It is outside the scope of this paper to go into the definition of the moment polytope. For that and a thorough discussion of all the connections to representation theory, invariant theory and quantum information theory we refer to [12]. Here we will discuss the properties of the quantum functionals that are important for upper bounding the asymptotic subrank. The first crucial fact proven in [12] is that the quantum functionals are in the asymptotic spectrum of all complex tensors $\Delta(\{\text{tensors over } \mathbb{C}\})$, which we discussed in Section 4.1. We conclude that we can lower bound irreversibility in terms of the flattening ranks and the quantum functionals as follows:

Proposition 6.1. *Let t be a tensor over the complex numbers. Then*

$$\mathbf{i}(t) \geq \frac{\max_i \log_2 \mathrm{R}(t_i)}{\min_\theta \log_2 F^\theta(t)}\,. \tag{6.1}$$

Comparing to Section 4.2, it is proved in [12] that the quantum functionals are at least as powerful as the support functionals when it comes to upper bounding the asymptotic subrank. Namely, for any tensor t over the complex numbers holds that $F^\theta(t) \leq \zeta^\theta(t)$. It is an open problem whether this inequality can be strict. We note that, as opposed to the support functionals, the quantum functionals are defined as convex programs. It is an open problem whether the quantum functionals are efficiently computable.

6.2 Asymptotic slice rank

We finish Section 5 by discussing the asymptotic slice rank as a method to upper bound asymptotic subrank, and the relations to the support functionals and the quantum functionals. The *slice rank* (introduced by Tao [33] in the context of the cap set problem) of a tensor t is the smallest number r such that t can be written as a sum of r slice rank one tensors. A *slice rank one tensor* is a tensor for which there is an $i \in [3]$ such that the flattening t_i has matrix rank one. The asymptotic subrank, the slice rank and the support functionals are related in the following way:

$$\underset{\sim}{\mathrm{Q}}(t) \leq \limsup_n \operatorname{slicerank}(t^{\otimes n})^{1/n} \leq \min_\theta \zeta^\theta(t)\,. \tag{6.2}$$

(See [12].) Thus, in an asymptotic fashion, the slice rank upper bounds the asymptotic subrank and hence lower bounds irreducibility. Any analysis of $\limsup_n \operatorname{slicerank}(t^{\otimes n})^{1/n}$ that we are aware of in the literature boils down to evaluating $\min_\theta \zeta^\theta(t)$. In particular, we are not aware of any example for which the right-most inequality in (6.2) is strict.

In fact, for oblique[4] tensors the right-most inequality is an equality [34] (see also [39, Section 4.6]) and, as mentioned in Section 4.2, for tight tensors both inequalities are equalities [31]. Over the complex numbers, the quantum functionals play a special role in this comparison. As we mentioned in Section 6.1, the quantum functionals satisfy $F^\theta(t) \leq \zeta^\theta(t)$. It was shown in [12] that the minimum over the parameter θ equals the asymptotic slice rank. We thus have

$$\underset{\sim}{\mathrm{Q}}(t) \leq \limsup_n \operatorname{slicerank}(t^{\otimes n})^{1/n} = \min_\theta F^\theta(t) \leq \min_\theta \zeta^\theta(t)\,. \tag{6.3}$$

For free[5] tensors the right-most inequality in (6.3) is an equality [12].

A Code to verify the numerical examples

The following Mathematica code generates the barrier values in the tables in Section 1 to arbitrary precision and for arbitrary parameters q and n.

A.1 Small Coppersmith–Winograd tensor cw_q

```
In[1]:= Table[
  2 Log2[q + 1]/(Log2[3] - 2/3 + (2/3) Log2[q]), {q, 2, 7}] // N

Out[1]= {2., 2.02538, 2.06244, 2.09627, 2.12549, 2.15064}
```

A.2 Big Coppersmith–Winograd tensor CW_q

```
In[1]:= g[b_, q_] := -((2 b q Log[2 b])/
   Log[2]) - ((1 - 3 b q) Log[1/3 (1 - 3 b q)])/(
  3 Log[2]) + ((-b q - 2/3 (1 - 3 b q)) Log[b q + 2/3 (1 - 3 b q)])/
  Log[2]

In[2]:= f[q_] :=
 Piecewise[{{g[
     q/(2 (-4 + q^2)) + 1/6 Sqrt[(32 + q^2)/(-4 + q^2)^2] /. {q -> 1},
      1], q == 1}, {g[1/9, q],
    q == 2}, {g[-((-3 q + Sqrt[32 + q^2])/(6 (-4 + q^2))), q],
    q >= 3}}]
```

[4] A tensor $t \in \mathbb{F}^{n_1} \otimes \mathbb{F}^{n_2} \otimes \mathbb{F}^{n_3}$ is called *oblique* if the support $\operatorname{supp}(t) \in [n_1] \times [n_2] \times [n_3]$ in some basis is an antichain in the product of the natural orders on the $[n_i]$. The matrix multiplication tensors $\langle a,b,c \rangle$ are examples of oblique tensors.

[5] A tensor t is called *free* if in some basis any two different $a,b \in \operatorname{supp}(t)$ differ in at least two entries. Every tight tensor is oblique and every oblique tensor is free.

```
In[3]:= Table[2 Log2[q + 2]/f[q], {q, 1, 6}] // N

Out[3]= {2.16805, 2.17795, 2.19146, 2.20551, 2.21913, 2.23201}
```

A.3 Reduced polynomial multiplication t_n

```
In[1]:= g[n_] := g[n] = g /. FindRoot[1/(g-1)-n/(g^n-1)==(n-1)/3,{g,1+2/n}]

In[2]:= z[n_] := (g[n]^n - 1)/(g[n] - 1) g[n]^(-2 (n - 1)/3)

In[3]:= Table[2 Log2[n]/Log2[z[n]], {n, 2, 7}]

Out[3]= {2.17795, 2.16805, 2.15949, 2.15237, 2.14641, 2.14135}
```

References

[1] JOSH ALMAN: Limits on the universal method for matrix multiplication. *Theory of Computing*, 17(1):1–30, 2021. Preliminary version in CCC'19. [doi:10.4086/toc.2021.v017a001, arXiv:1812.08731] 3, 5

[2] JOSH ALMAN AND VIRGINIA VASSILEVSKA WILLIAMS: Further limitations of the known approaches for matrix multiplication. In *Proc. 9th Innovations in Theoret. Comp. Sci. conf. (ITCS'18)*, pp. 25:1–15. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018. [doi:10.4230/LIPIcs.ITCS.2018.25] 2

[3] JOSH ALMAN AND VIRGINIA VASSILEVSKA WILLIAMS: Limits on all known (and some unknown) approaches to matrix multiplication. In *Proc. 59th FOCS*, pp. 580–591. IEEE Comp. Soc., 2018. [doi:10.1109/FOCS.2018.00061, arXiv:1810.08671] 2, 5

[4] ANDRIS AMBAINIS, YUVAL FILMUS, AND FRANÇOIS LE GALL: Fast matrix multiplication: limitations of the Coppersmith–Winograd method (extended abstract). In *Proc. 47th STOC*, pp. 585–593. ACM Press, 2015. [doi:10.1145/2746539.2746554] 2, 5

[5] FELIX A. BEHREND: On sets of integers which contain no three terms in arithmetical progression. *Proc. Nat. Acad. Sci. USA*, 32(12):331–332, 1946. [doi:10.1073/pnas.32.12.331] 8

[6] CHARLES H. BENNETT, SANDU POPESCU, DANIEL ROHRLICH, JOHN A. SMOLIN, AND ASHISH V. THAPLIYAL: Exact and asymptotic measures of multipartite pure-state entanglement. *Phys. Rev. A (3)*, 63(1):012307, 2000. [doi:10.1103/PhysRevA.63.012307] 5

[7] MARKUS BLÄSER: *Fast Matrix Multiplication*. Number 5 in Graduate Surveys. Theory of Computing Library, 2013. [doi:10.4086/toc.gs.2013.005] 5, 7, 22

[8] JONAH BLASIAK, THOMAS CHURCH, HENRY COHN, JOSHUA A. GROCHOW, ERIC NASLUND, WILLIAM F. SAWIN, AND CHRIS UMANS: On cap sets and the group-theoretic approach to matrix multiplication. *Discrete Analysis*, 3:1–27, 2017. [doi:10.19086/da.1245, arXiv:1605.06702] 2, 5, 12, 24, 25

[9] JONAH BLASIAK, THOMAS CHURCH, HENRY COHN, JOSHUA A. GROCHOW, AND CHRIS UMANS: Which groups are amenable to proving exponent two for matrix multiplication?, 2017. [arXiv:1712.02302] 2, 5, 25

[10] PETER BÜRGISSER, MICHAEL CLAUSEN, AND M. AMIN SHOKROLLAHI: *Algebraic Complexity Theory*. Volume 315 of *Grundlehren der Math. Wiss.* Springer, 1997. [doi:10.1007/978-3-662-03338-8] 5, 13

[11] MATTHIAS CHRISTANDL, FRANÇOIS LE GALL, VLADIMIR LYSIKOV, AND JEROEN ZUIDDAM: Barriers for rectangular matrix multiplication, 2020. [arXiv:2003.03019] 5

[12] MATTHIAS CHRISTANDL, PÉTER VRANA, AND JEROEN ZUIDDAM: Universal points in the asymptotic spectrum of tensors. In *Proc. 50th STOC*, pp. 289–296. ACM Press, 2018. [doi:10.1145/3188745.3188766, arXiv:1709.07851] 3, 4, 12, 21, 26, 27

[13] MATTHIAS CHRISTANDL, PÉTER VRANA, AND JEROEN ZUIDDAM: Asymptotic tensor rank of graph tensors: Beyond matrix multiplication. *Comput. Complexity*, 28(1):57–111, 2019. [doi:10.1007/s00037-018-0172-8, arXiv:1609.07476] 9

[14] MATTHIAS CHRISTANDL, PÉTER VRANA, AND JEROEN ZUIDDAM: Barriers for fast matrix multiplication from irreversibility. In *Proc. 34th Comput. Complexity Conf. (CCC'19)*, pp. 26:1–17. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019. [doi:10.4230/LIPIcs.CCC.2019.26] 1

[15] HENRY COHN AND CHRISTOPHER UMANS: A group-theoretic approach to fast matrix multiplication. In *Proc. 44th FOCS*, pp. 438–449. IEEE Comp. Soc., 2003. [doi:10.1109/SFCS.2003.1238217] 2, 25

[16] HENRY COHN AND CHRISTOPHER UMANS: Fast matrix multiplication using coherent configurations. In *Proc. 24th Ann. ACM–SIAM Symp. on Discrete Algorithms (SODA'13)*, pp. 1074–1086. SIAM, 2013. ACM DL. 2

[17] DON COPPERSMITH AND SHMUEL WINOGRAD: Matrix multiplication via arithmetic progressions. *J. Symbolic Comput.*, 9(3):251–280, 1990. [doi:10.1016/S0747-7171(08)80013-2] 2, 4, 21, 22, 23

[18] WOLFGANG DÜR, GUIVRE VIDAL, AND JUAN IGNACIO CIRAC: Three qubits can be entangled in two inequivalent ways. *Phys. Rev. A (3)*, 62(6):062314, 12, 2000. [doi:10.1103/PhysRevA.62.062314] 5

[19] JORDAN S. ELLENBERG AND DION GIJSWIJT: On large subsets of $\mathbf{F}_q^n$ with no three-term arithmetic progression. *Ann. Math.*, 185(1):339–343, 2017. [doi:10.4007/annals.2017.185.1.8] 12

[20] RYSZARD HORODECKI, PAWEŁ HORODECKI, MICHAŁ HORODECKI, AND KAROL HORODECKI: Quantum entanglement. *Rev. Modern Physics*, 81(2):865–942, 2009. [doi:10.1103/RevModPhys.81.865] 5

[21] SWASTIK KOPPARTY, GUY MOSHKOVITZ, AND JEROEN ZUIDDAM: Geometric rank of tensors and subrank of matrix multiplication. In *Proc. 35th Comput. Complexity Conf. (CCC'20)*, pp. 35:1–19. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2020. [doi:10.4230/LIPIcs.CCC.2020.35, arXiv:2002.09472] 8

[22] JOSEPH M. LANDSBERG AND MATEUSZ MICHAŁEK: Abelian tensors. *J. Math. Pures Appliquées*, 108(3):333–371, 2017. [doi:10.1016/j.matpur.2016.11.004, arXiv:1504.03732] 13

[23] FRANÇOIS LE GALL: Powers of tensors and fast matrix multiplication. In *Proc. 39th Internat. Symp. Symbolic and Algebraic Computation (ISSAC'14)*, pp. 296–303. ACM Press, 2014. [doi:10.1145/2608628.2608664] 2

[24] GEORGE PÓLYA AND GABOR SZEGŐ: *Problems and Theorems in Analysis I*. Classics in Math. Springer, 1924/1972/1998. [doi:10.1007/978-3-642-61983-0] 6

[25] RAPHAËL SALEM AND DONALD CLAYTON SPENCER: On sets of integers which contain no three terms in arithmetical progression. *Proc. Nat. Acad. Sci. USA*, 28(12):561–563, 1942. [doi:10.1073/pnas.28.12.561] 8

[26] WILL SAWIN: Bounds for matchings in nonabelian groups. *Electr. J. Combinat.*, 25(4), 2018. [doi:10.37236/7520, arXiv:1702.00905] 25

[27] ANDREW JAMES STOTHERS: *On the complexity of matrix multiplication*. Ph. D. thesis, Univ. Edinburgh, 2010. era-handle. 2

[28] VOLKER STRASSEN: Gaussian elimination is not optimal. *Numerische Mathematik*, 13(4):354–356, 1969. [doi:10.1007/BF02165411] 2

[29] VOLKER STRASSEN: Relative bilinear complexity and matrix multiplication. *J. reine angew. Math.*, 375/376:406–443, 1987. [doi:10.1515/crll.1987.375-376.406] 3, 10, 11

[30] VOLKER STRASSEN: The asymptotic spectrum of tensors. *J. reine angew. Math.*, 384:102–152, 1988. [doi:10.1515/crll.1988.384.102] 3, 7, 8, 12, 13, 15, 18

[31] VOLKER STRASSEN: Degeneration and complexity of bilinear maps: Some asymptotic spectra. *J. reine angew. Math.*, 413:127–180, 1991. [doi:10.1515/crll.1991.413.127] 3, 4, 10, 12, 19, 20, 21, 24, 27

[32] VOLKER STRASSEN: Algebra and complexity. In *First Eur. Congr. Math., Vol. II*, volume 120 of *Progress Math.*, pp. 429–446. Birkhäuser, 1994. [doi:10.1007/978-3-0348-9112-7_18] 13

[33] TERENCE TAO: A symmetric formulation of the Croot–Lev–Pach–Ellenberg–Gijswijt capset bound, 2016. LINK at terrytao.wordpress.com. 12, 26

[34] TERENCE TAO AND WILL SAWIN: Notes on the “slice rank” of tensors, 2016. LINK at terrytao.wordpress.com. 25, 27

[35] VIRGINIA VASSILEVSKA WILLIAMS: Multiplying matrices faster than Coppersmith–Winograd. Extended abstract. In *Proc. 44th STOC*, pp. 887–898. ACM Press, 2012. [doi:10.1145/2213977.2214056] 2

[36] FRANK VERSTRAETE, JEROEN DEHAENE, BART L. R. DE MOOR, AND HENRI VERSCHELDE: Four qubits can be entangled in nine different ways. *Phys. Rev. A (3)*, 65(5):052112, 5, 2002. [doi:10.1103/PhysRevA.65.052112] 5

[37] PÉTER VRANA AND MATTHIAS CHRISTANDL: Asymptotic entanglement transformation between W and GHZ states. *J. Math. Phys.*, 56(2):022204, 12, 2015. [doi:10.1063/1.4908106] 9

[38] NENGKUN YU, CHENG GUO, AND RUNYAO DUAN: Obtaining a W state from a Greenberger–Horne–Zeilinger state via stochastic local operations and classical communication with a rate approaching unity. *Phys. Rev. Lett.*, 112(16):160401, 2014. [doi:10.1103/PhysRevLett.112.160401] 9

[39] JEROEN ZUIDDAM: *Algebraic complexity, asymptotic spectra and entanglement polytopes*. Ph. D. thesis, Univ. Amsterdam, 2018. UvA-DARE. 27

AUTHORS

Matthias Christandl
Professor
Department of Mathematical Sciences
University of Copenhagen
Copenhagen, Denmark
christandl@math.ku.dk

Péter Vrana
Associate professor
Department of Geometry
Budapest University of Technology and Economics
Budapest, Hungary

MTA-BME Lendület Quantum Information Theory Research Group

vranap@math.bme.hu
http://math.bme.hu/~vranap/

Jeroen Zuiddam
Assistant professor
Korteweg–de Vries Institute for Mathematics
University of Amsterdam
Amsterdam, Netherlands
j.zuiddam@uva.nl
https://staff.fnwi.uva.nl/j.zuiddam/

ABOUT THE AUTHORS

MATTHIAS CHRISTANDL is a professor at the Department of Mathematical Sciences at the University of Copenhagen in Denmark. He received his Ph. D. from the University of Cambridge in 2006. Prior to joining Copenhagen in 2014 he was an assistant professor at ETH Zurich. His research interests are in quantum information theory, quantum computation and algebraic complexity theory.

PÉTER VRANA received his Ph. D. in 2011 from the Doctoral School of Physical Sciences, Budapest University of Technology and Economics, under the supervision of Péter Lévay. His main research interest is in quantum information theory. Péter lives in Budapest, Hungary, where he has been a faculty member of the BUTE Institute of Mathematics since 2014.

JEROEN ZUIDDAM received his Ph. D. from the University of Amsterdam in 2018, under the supervision of Harry Buhrman and Matthias Christandl. During that time he was a member of the Centrum Wiskunde & Informatica and the Institute for Logic, Language and Computation. From 2018 to 2020, Jeroen was a member of the Institute for Advanced Study in Princeton, where he worked with Avi Wigderson. From 2020 to 2021, Jeroen was a Simons Junior Fellow at the Courant Institute of Mathematical Sciences of New York University, hosted by Oded Regev. He is currently an assistant professor at the Korteweg–de Vries Institute for Mathematics at the University of Amsterdam. His research ranges over several areas of computer science and mathematics, including algebraic complexity theory, quantum information theory and discrete mathematics.