Asymptotic tensor rank of graph tensors
Christandl, Matthias; Vrana, Péter; Zuiddam, Jeroen

Published in:
Computational Complexity

DOI:
10.1007/s00037-018-0172-8

Publication date:
2019

Document version
Peer reviewed version

Citation for published version (APA):
Asymptotic tensor rank of graph tensors: beyond matrix multiplication

Matthias Christandl\textsuperscript{1}, Péter Vrana\textsuperscript{2} and Jeroen Zuiddam\textsuperscript{3}

Abstract

We present an upper bound on the exponent of the asymptotic behaviour of the tensor rank of a family of tensors defined by the complete graph on $k$ vertices. For $k \geq 4$, we show that the exponent per edge is at most 0.77, outperforming the best known upper bound on the exponent per edge for matrix multiplication ($k = 3$), which is approximately 0.79. We raise the question whether for some $k$ the exponent per edge can be below $\frac{2}{3}$, i.e. can outperform matrix multiplication even if the matrix multiplication exponent equals 2. In order to obtain our results, we generalise to higher order tensors a result by Strassen on the asymptotic subrank of tight tensors and a result by Coppersmith and Winograd on the asymptotic rank of matrix multiplication. Our results have applications in entanglement theory and communication complexity.

1. Introduction

A famous open problem in algebraic complexity theory is the problem of determining the tensor rank of large tensor powers of the $2 \times 2$ matrix multiplication tensor. Let $V_1, \ldots, V_k$ be finite-dimensional complex vector spaces and let $\phi \in V_1 \otimes \cdots \otimes V_k$ be a $k$-tensor. The tensor rank $R(\phi)$ of $\phi$ is the smallest number $r$ such that $\phi$ can be written as a sum of $r$ simple tensors $v_1 \otimes \cdots \otimes v_k \in V_1 \otimes \cdots \otimes V_k$. We define the $N$th tensor power $\phi^{\otimes N}$ as the $k$-tensor obtained by taking the tensor product of $N$ copies of $\phi$ and grouping such that $\phi \in (V_1^{\otimes N}) \otimes \cdots \otimes (V_k^{\otimes N})$. For any $n$, let $b_1, \ldots, b_n$ denote the standard basis of $\mathbb{C}^n$. The $2 \times 2$ matrix multiplication tensor is the 3-tensor

$$\sum_{i \in \{0,1\}^3} (b_{i_1} \otimes b_{i_2}) \otimes (b_{i_2} \otimes b_{i_3}) \otimes (b_{i_3} \otimes b_{i_1}) \in (\mathbb{C}^2 \otimes \mathbb{C}^2)^{\otimes 3}.$$
This paper is motivated by the study of the tensor rank of powers of tensors that are generalizations of the matrix multiplication tensor. Consider a graph with two vertices connected by a single edge. We define the corresponding 2-tensor as

$$T(\leftrightarrow) = \sum_{i \in \{0,1\}} b_i \otimes b_i \in \mathbb{C}^2 \otimes \mathbb{C}^2.$$  

Let $G$ be any graph. Then we define $T(G)$ to be the $|V|$-tensor obtained by taking the tensor product of the tensors corresponding to the edges of $G$, grouping corresponding vertices together (the full definition is in Theorem 1.1.1). For example, for the complete graph on four vertices $K_4$ we have

$$T(K_4) = T(\leftrightarrow) \otimes T(\leftrightarrow) \otimes T(\leftrightarrow) \otimes T(\leftrightarrow) \otimes T(\leftrightarrow) \otimes T(\leftrightarrow)$$

$$= \sum_{i \in \{0,1\}^6} (b_{i_1} \otimes b_{i_2} \otimes b_{i_3}) \otimes (b_{i_4} \otimes b_{i_5} \otimes b_{i_6}) \otimes (b_{i_7} \otimes b_{i_8} \otimes b_{i_9})$$

living in $(\mathbb{C}^8)^\otimes 4$. We can ignore the dependence of this tensor on the order of the edges, since tensor rank is invariant under this choice. Let $K_k$ be the complete graph on $k$ vertices. The $2 \times 2$ matrix multiplication tensor is the tensor $T(K_3)$.

The main result of this paper is an upper bound on the tensor rank of large tensor powers of $T(K_k)$: for any $k \geq 4$ and large $N$,

$$R(T(K_k)^\otimes N) \leq 2^{0.772943(k/2)^N + o(N)}.$$  

We say that the exponent $\omega(T(K_k))$ is at most $0.772943(k/2)$ and we say that the exponent per edge $\tau(T(K_k))$ is at most $0.772943$. This improves, for $k \geq 4$, the bound $\tau(T(K_k)) \leq 0.790955$ that can be derived from the well-known upper bound of Le Gall [LG14] on the exponent of matrix multiplication $\omega := \omega(T(K_3))$. Note that $\tau(T(K_3)) = \omega/3$.

By a “covering argument” we can show that $\tau(T(K_k))$ is nonincreasing when $k$ increases (Theorem 1.1.25). On the other hand, a standard “flattening argument” (see Theorem 1.1.19) yields the lower bound $\tau(T(K_k)) \geq \frac{1}{2}k/(k-1)$ if $k$ is even and $\tau(T(K_k)) \geq \frac{1}{2}(k+1)/k$ if $k$ is odd. As a consequence, if the exponent of matrix multiplication $\omega$ equals 2, then $\tau(T(K_4)) = \tau(T(K_3)) = \frac{2}{3}$. We raise the following question: is there a $k \geq 5$ such that $\tau(T(K_k)) < \frac{2}{3}$? More open questions are discussed in Section 1.3.

Our method to prove (1) is a generalization of a method of Strassen, and Coppersmith and Winograd for obtaining upper bounds on the exponent of the matrix multiplication tensor. To use the generalized Coppersmith–Winograd method we have to get a handle on the monomial subrank of tensor powers of yet another type of combinatorially defined tensors. The definition of rank given above is equivalent to saying that the rank $R(\phi)$ of a tensor
We are interested in lower bounding \( T \), which there is a choice of bases where \( y \) yields the monomial subrank \( Q(\phi) \) of \( \phi \) is the largest number \( s \) for which there exist linear maps \( A_1 : \mathbb{C}^r \to V_1 \), \ldots, \( A_k : \mathbb{C}^r \to V_k \) such that \( \phi = (A_1 \otimes \cdots \otimes A_k) T_r(k) \), where
\[
T_r(k) = \sum_{i=1}^r (b_i)^{\otimes k}
\]
is the rank-\( r \) unit \( k \)-tensor. The subrank \( Q(\phi) \) of \( \phi \) is the largest number \( s \) for which there exist linear maps \( A_1 : V_1 \to \mathbb{C}^s \), \ldots, \( A_k : V_k \to \mathbb{C}^s \) such that \( T_s(k) = (A_1 \otimes \cdots \otimes A_k) \phi \). In fact, we need a slightly restricted notion. A monomial matrix is a matrix such that any row or column has at most one nonzero entry. Let \( \phi \in V_1 \otimes \cdots \otimes V_k \) be a tensor in a fixed basis. The monomial subrank \( Q_M(\phi) \) is the largest number \( s \) for which there exist monomial matrices \( A_1, \ldots, A_k \) such that \( T_s(k) = (A_1 \otimes \cdots \otimes A_k) \phi \).

We are interested in lower bounding \( Q_M(D_{(2,2)}^{\otimes N}) \) for large \( N \). In particular, in the course of proving (1) we face the problem of computing \( Q_M(D_{(2,2)}^{\otimes N}) \), where
\[
D_{(2,2)} = b_1 \otimes b_1 \otimes b_2 \otimes b_2 + b_1 \otimes b_2 \otimes b_1 \otimes b_2 + b_1 \otimes b_2 \otimes b_2 \otimes b_1 \\
+ b_2 \otimes b_1 \otimes b_1 \otimes b_2 + b_2 \otimes b_1 \otimes b_2 \otimes b_1 + b_2 \otimes b_2 \otimes b_1 \otimes b_1
\]
is the weight-(2,2) Dicke tensor.

Our second result solves this problem. Namely, we prove a general asymptotic lower bound on the monomial subrank of tensor powers of so-called tight tensors (Theorem 1.2.4). These are tensors \( \phi \in V_1 \otimes \cdots \otimes V_k \) for which there is a choice of bases \( B_1, \ldots, B_k \) for \( V_1, \ldots, V_k \) respectively and injective maps \( \alpha_1 : B_1 \to \mathbb{Z} \), \ldots, \( \alpha_k : B_k \to \mathbb{Z} \) such that
\[
\forall (b_1, \ldots, b_k) \in \text{supp}_B(\phi) \quad \alpha_1(b_1) + \cdots + \alpha_k(b_k) = 0.
\]
For example, applied to the tensor \( D_{(2,2)} \), which is a tight tensor, our result yields the monomial subrank \( Q_M(D_{(2,2)}^{\otimes N}) = 2^N - o(N) \), which is asymptotically optimal. We say that the monomial subexponent \( q_M(D_{(2,2)}) \) equals 1. This solves a conjecture posed in [VC15] for the special case \( D_{(2,2)} \).

In quantum information theoretical terms, the tensor \( T(K_k) \) encodes a \( k \)-partite quantum state in which each two systems share an Einstein–Podolski–Rosen (EPR) pair. The tensor \( T_2(k) \) encodes the \( k \)-partite Greenberger–Horne–Zeilinger (GHZ) state. If the rank of \( T(K_k) \) is \( r \), then \( \lceil \log_2 r \rceil \) copies of the state \( T_2(k) \) are needed to generate \( T(K_k) \) by stochastic local operations and classical communication (SLOCC). See [DVC00] for a discussion of SLOCC. Asymptotically, we can generate \( N \) copies of \( T(K_k) \) using \( \omega(T(K_k)) N + o(N) \) copies of \( T(k) \).

In the next subsection we discuss preliminary definitions and known results. After that we state our results and discuss open questions.

### 1.1. Preliminaries

Let \( U_1, \ldots, U_k \) and \( V_1, \ldots, V_k \) be complex finite-dimensional vector spaces. For any number \( n \in \mathbb{N} \), define the set \( [n] := \{1, 2, \ldots, n\} \). All our graphs will
be simple graphs, that is, they are unweighted, undirected, containing no self-loops or multiple edges.

We first define the two families of tensors that play an important role in this paper, namely the tensors $T(G)$ with $G$ a graph and the tensors $D_\lambda$ with $\lambda$ a partition. We start with $T(G)$.

**Definition 1.1.1.** Let $G = (V, E)$ be a graph and let $n$ be a natural number. Let $b_1, \ldots, b_n$ be the standard basis of $\mathbb{C}^n$. We define the $|V|$-tensor $T_n(G)$ as

$$T_n(G) := \sum_{i \in [n]^E} \bigotimes_{e \in V; v \in e} b_{i_e},$$

where the sum is over all tuples $i$ indexed by $E$ with entries in $[n]$. Equivalently, one can define $T_n(G)$ as a tensor product over edges of $G$, as follows, letting subscripts denote the position of tensor legs,

$$T_n(G) = \bigotimes_{e \in E} \sum_{i \in [n]} (b_{i_1} \otimes b_{i_2})_e \otimes (1 \otimes \cdots \otimes 1)_{V \setminus e}.$$ 

Here, the element $1$ should be regarded as an element of $\mathbb{C}$ and the large tensor product inputs and outputs $|V|$-tensors by the natural regrouping. We will denote $T_2(G)$ by $T(G)$. The definition naturally generalizes to hypergraphs, see [VC17].

Note that the tensor $T_n(G)$ behaves as follows under the tensor Kronecker product: $T_n(G) \otimes T_m(G) \cong T_{nm}(G)$.

We can ignore the fact that the tensor in the above definition depends on the choice of order of the edges and vertices of the graph $G$, since tensor rank and subrank do not depend on this order.

Let $K_k$ be the complete graph on $k$ vertices and let $C_k$ be the cycle graph on $k$ vertices.

**Example 1.1.2.** We give some examples of tensors of type $T_n(G)$. Let $n \in \mathbb{N}$. For any $j_1, j_2, j_3 \in [n]$, we define $b_{j_1j_2} := b_{j_1} \otimes b_{j_2}$ and $b_{j_1j_2j_3} := b_{j_1} \otimes b_{j_2} \otimes b_{j_3}$. Then

$$T_n(K_2) = \sum_{i \in [n]^2} b_{i_1} \otimes b_{i_2} \in \mathbb{C}^n \otimes \mathbb{C}^n$$

$$T_n(K_3) = \sum_{i \in [n]^3} b_{i_1i_2} \otimes b_{i_2i_3} \otimes b_{i_3i_1} \in \mathbb{C}^{n^2} \otimes \mathbb{C}^{n^2} \otimes \mathbb{C}^{n^2}$$

$$T_n(K_4) = \sum_{i \in [n]^6} b_{i_1i_2i_3} \otimes b_{i_2i_3i_4} \otimes b_{i_3i_4i_5} \otimes b_{i_4i_5i_1} \in (\mathbb{C}^{n^3})^\otimes 4$$

$$T_n(C_5) = \sum_{i \in [n]^5} b_{i_1i_2} \otimes b_{i_2i_3} \otimes b_{i_3i_4} \otimes b_{i_4i_5} \otimes b_{i_5i_1} \in (\mathbb{C}^{n^2})^\otimes 5$$


In algebraic complexity theory, the tensor $T_n(C_3)$ is called the $n \times n$ matrix multiplication tensor, since it encodes the bilinear map that multiplies two $n \times n$ matrices. It is usually denoted by $\langle n,n,n \rangle$. For $k \geq 3$, $T_n(C_k)$ is the iterated matrix multiplication tensor; it encodes the multilinear map that multiplies $k$ matrices of size $n \times n$. Note that $C_3 = K_3$, so any results on $T_n(K_k)$ or $T_n(C_k)$ are generalizations of results on the matrix multiplication tensor.

We now introduce the second family of tensors, $D_\lambda$.

**Definition 1.1.3** (Dicke tensor [Dic54, SGDM03, VC17]). Let $k$ be a positive integer and let $\lambda = (\lambda_1, \ldots, \lambda_n) \vdash n$ be a partition of $k$ of length at most $n$. Let $b_1, \ldots, b_n$ be the standard basis of $\mathbb{C}^n$. Define the weight-$\lambda$ Dicke tensor $D_\lambda$ as

$$D_\lambda := \sum_{i \in [n]^k : \text{type}(i) = \lambda} b_{i_1} \otimes \cdots \otimes b_{i_k} \in (\mathbb{C}^n)^\otimes k$$

where type$(i) = \lambda$ means that $i$ is a permutation of the tuple $(\underbrace{1, \ldots, 1}_\lambda, \underbrace{2, \ldots, 2}_\lambda, \ldots, \underbrace{n, \ldots, n}_\lambda)$.

**Example 1.1.4.** We give some examples of tensors of type $D_\lambda$.

$$D_{(2,1)} = b_1 \otimes b_1 \otimes b_2 + b_1 \otimes b_2 \otimes b_1 + b_2 \otimes b_1 \otimes b_1,$$

$$D_{(1,1,1)} = b_1 \otimes b_2 \otimes b_3 + b_1 \otimes b_3 \otimes b_2 + b_2 \otimes b_1 \otimes b_3$$

$$+ b_2 \otimes b_3 \otimes b_1 + b_3 \otimes b_1 \otimes b_2 + b_3 \otimes b_2 \otimes b_1.$$

In quantum information theory, the tensor $D_{(2,1)}$ encodes a quantum state known as the W-state, and $D_{(k-1,1)}$ encodes a generalized W-state on $k$ systems.

We now wish to define the exponent and (monomial) subexponent of a tensor, which should be thought of as our complexity measures for tensors. First we define (monomial) restriction, degeneration and asymptotic conversion rate.

Let $\phi \in U_1 \otimes \cdots \otimes U_k$ and $\psi \in V_1 \otimes \cdots \otimes V_k$ be $k$-tensors.

**Definition 1.1.5.** We say $\phi$ restricts to $\psi$, written $\phi \geq \psi$, if there exist linear maps $A_i : U_i \to V_i$ such that $\psi = (A_1 \otimes \cdots \otimes A_k)\phi$. We say $\phi$ is isomorphic to $\psi$ if both $\phi \geq \psi$ and $\phi \leq \psi$. We will often tacitly treat isomorphic tensors as being equal. We say $\phi$ degenerates to $\psi$, written $\phi \succeq \psi$, if $\psi$ is in the orbit closure $\overline{G\phi}$ in the Zariski topology, where $G = GL(U_1) \times \cdots \times GL(U_k)$.

A matrix is a monomial matrix if on any of its rows or columns there is at most one nonzero entry. Fix bases for $U_1, \ldots, U_k$ and $V_1, \ldots, V_k$. We
say that there is a \textit{monomial restriction} from $\phi$ to $\psi$, written $\phi \geq_M \psi$, if there exist monomial matrices $A_1, \ldots, A_k$ such that $\psi = (A_1 \otimes \cdots \otimes A_k) \phi$ in the chosen basis. We say that there is a \textit{monomial degeneration} from $\phi$ to $\psi$, written $\phi \vM \psi$, if $\psi$ is in the orbit closure $M\phi$ in the Zariski topology, where $M$ is the subgroup of $G$ consisting of $k$-tuples of monomial matrices.

Monomial degeneration has a nice combinatorial description for which we refer to Theorem 6.1 in \[Str87\]. It is clear from the definition that $\phi \geq_M \psi$ implies $\phi \vM \psi$. We refer to \[Str87\] for other basic properties of $\geq_M$ and $\vM$.

\textbf{Definition 1.1.6.} Define the \textit{asymptotic conversion rate} from $\phi$ to $\psi$ as

$$\omega(\phi, \psi) := \lim_{n \to \infty} \frac{1}{n} \min \{ m \in \mathbb{N} \mid \phi^\otimes m \geq \psi^\otimes n \},$$

and the \textit{asymptotic monomial conversion rate} from $\phi$ to $\psi$ as

$$\omega_M(\phi, \psi) := \lim_{n \to \infty} \frac{1}{n} \min \{ m \in \mathbb{N} \mid \phi^\otimes m \vM \psi^\otimes n \}.$$  

The minimum of the empty set is considered to be $\infty$.

\textbf{Proposition 1.1.7.} The limit in Equation (2) exists and equals the supremum $\sup_n \frac{1}{n} \min \{ m \in \mathbb{N} \mid \phi^\otimes m \geq \psi^\otimes n \}$. The limit in Equation (3) exists and equals the supremum $\sup_n \frac{1}{n} \min \{ m \in \mathbb{N} \mid \phi^\otimes m \vM \psi^\otimes n \}$.

\textit{Proof.} See Lemma 1.1 in \[Str88\]. \qed

\textbf{Theorem 1.1.8.} Restriction and degeneration are asymptotically equivalent, in the sense that

$$\omega(\phi, \psi) = \lim_{n \to \infty} \frac{1}{n} \min \{ m \in \mathbb{N} \mid \phi^\otimes m \geq \psi^\otimes n \}.$$  

\textit{Proof.} See \[Str87\]. A proof can also be found in \[VC15\]. \qed

\textbf{Remark 1.1.9.} We do not know whether in (3) one may equivalently replace $\vM$ by $\geq_M$. We therefore defined $\omega_M$ using the more powerful $\vM$.

Strassen introduced the asymptotic study of restriction in the context of the algebraic complexity of bilinear maps \[Str88, Str91\]. In quantum information theory, if $\phi$ and $\psi$ are pure quantum states, then $\phi \geq \psi$ means precisely that $\psi$ can be obtained from $\phi$ by means of stochastic local operations and classical communication (SLOCC). We refer to \[VC15\] and \[VC17\] for general properties of $\omega(\phi, \psi)$.

Restriction and asymptotic conversion rate can be used to compare any two $k$-tensors. In this context, the following tensor will serve as an absolute reference tensor.
Definition 1.1.10 (GHZ tensor or unit tensor). Let $k, r \in \mathbb{N}$, and let $b_1, \ldots, b_r$ be the standard basis of $\mathbb{C}^r$. Define $T_r(k)$ as the $k$-tensor

$$T_r(k) := \sum_{i \in [r]} b_i \otimes \cdots \otimes b_i \in (\mathbb{C}^r)^\otimes k.$$ 

We will denote $T_r(k)$ by $T_r$ when the order $k$ is understood and we will denote $T_2(k)$ by just $T(k)$ or $T$. In quantum information theory, the tensor $T_r(k)$ encodes the Greenberger–Horne–Zeilinger (GHZ) quantum state of rank $r$ on $k$ systems. In algebraic complexity theory, for $k = 3$ this tensor is called the rank-$r$ unit tensor and is denoted by $\langle r \rangle$. Note that $T_r(k)$ equals the hypergraph tensor $T_r(H)$ where $H$ is the hypergraph on $k$ vertices with a single hyperedge containing all vertices. We refer to [VC17] for the definition of hypergraph tensor.

Definition 1.1.11. The \textit{rank} of $\psi$, denoted $R(\psi)$, is the smallest number $r$ such that $\psi \leq T_r$. Equivalently, it is the smallest number $r$ such that $\psi$ can be written as a sum of $r$ simple tensors $v_1 \otimes \cdots \otimes v_k \in V_1 \otimes \cdots \otimes V_k$. The \textit{border rank} of $\psi$, denoted $R(\psi)$ is the smallest number $r$ such that $\psi \leq T_r$.

Definition 1.1.12. The \textit{subrank} of $\phi$, denoted $Q(\phi)$, is the largest number $s$ such that $T_s \leq \phi$. The \textit{monomial subrank} of $\phi$, denoted $Q_M(\phi)$, is the largest number $s$ such that $T_s \leq_M \phi$. In the same way, one defines the \textit{border subrank} $Q(\phi)$ and the \textit{monomial border subrank} $Q_M(\phi)$ using $\leq$ and $\leq_M$ respectively.

Definition 1.1.13. The \textit{exponent} of $\psi$ is defined as the asymptotic conversion rate from $T$ to $\psi$ and is denoted by $\omega(\psi)$,

$$\omega(\psi) := \omega(T, \psi).$$

Definition 1.1.14. The \textit{subexponent} of $\phi$ is defined as the inverse of the asymptotic conversion rate from $\psi$ to $T$, and is denoted by $q(\phi)$,

$$q(\phi) := \omega(\phi, T)^{-1}.$$ 

Definition 1.1.15. Let $\phi \in U_1 \otimes \cdots \otimes U_k$ be a $k$-tensor in a fixed basis. The \textit{monomial subexponent} of $\phi$ is defined as the inverse of the asymptotic monomial conversion rate from $\psi$ to $T$, and is denoted by $q_M(\phi)$,

$$q_M(\phi) := \omega_M(\phi, T)^{-1}.$$ 

The parameters $\omega$, $q$ and $q_M$ have the following two useful descriptions in terms of $R$, $Q$ and $Q_M$.

Proposition 1.1.16. Let $\phi$ and $\psi$ be $k$-tensors. Then,

\begin{align*}
\omega(\psi) &= \lim_{n \to \infty} \frac{1}{n} \log_2 R(\psi^\otimes n), \\
q(\phi) &= \lim_{m \to \infty} \frac{1}{m} \log_2 Q(\phi^\otimes m), \\
q_M(\phi) &= \lim_{m \to \infty} \frac{1}{m} \log_2 Q_M(\phi^\otimes m).
\end{align*}
Proof. To prove the first equality, we have
\[
\omega(T, \psi) = \lim_{n \to \infty} \frac{1}{n} \min \{ m \in \mathbb{N} \mid T^\otimes m \geq \psi^\otimes n \} \\
= \lim_{n \to \infty} \frac{1}{n} \left\lceil \log_2 R(\psi^\otimes n) \right\rceil \\
= \lim_{n \to \infty} \frac{1}{n} \log_2 R(\psi^\otimes n).
\]
To prove the second equality, we have
\[
\omega(\phi, T) = \lim_{n \to \infty} \frac{1}{n} \min \{ m \in \mathbb{N} \mid \phi^\otimes m \geq T^\otimes n \} \\
= \lim_{m \to \infty} \max \{ n \in \mathbb{N} \mid \phi^\otimes m \geq T^\otimes n \}^{-1} m \\
= \lim_{m \to \infty} \left( \log_2 Q(\phi^\otimes M) \right)^{-1} m \\
= \lim_{m \to \infty} \left( \log_2 Q(\phi^\otimes M) \right)^{-1} m \\
= q(\phi)^{-1}.
\]
The statement for \( q_M \) follows from a similar proof.

Proposition 1.1.17. Let \( \phi \) and \( \psi \) be \( k \)-tensors. Then,
\[
\omega(\psi) = \inf \{ \beta \in \mathbb{R} \mid R(\psi^\otimes N) \leq 2^{\beta N + o(N)} \},
q(\phi) = \sup \{ \beta \in \mathbb{R} \mid Q(\phi^\otimes N) \geq 2^{\beta N - o(N)} \},
q_M(\phi) = \sup \{ \beta \in \mathbb{R} \mid Q_M(\phi^\otimes N) \geq 2^{\beta N - o(N)} \}.
\]
Proof. If \( R(\psi^\otimes N) \leq 2^{\beta N + o(N)} \), then \( \omega(\psi) \leq \lim_{N \to \infty} \beta + o(N) / N = \beta \), by Theorem 1.1.16. Conversely, suppose that \( \omega(\psi) < \beta \). Then \( R(\psi^\otimes N) \leq O(2^{\beta N}) \), by Theorem 1.1.16, so \( R(\psi^\otimes N) \leq 2^{\beta N + o(N)} \). The statement for \( q \) and \( q_M \) follows similarly from Theorem 1.1.16.

For tensors of type \( T(G) \) we have the following characterization of \( \omega \), \( q \) and \( q_M \).

Proposition 1.1.18. Let \( G \) be a graph. Then,
\[
\omega(T(G)) = \inf \{ \beta \in \mathbb{R} \mid R(T_n(G)) = O(n^\beta) \},
q(T(G)) = \sup \{ \beta \in \mathbb{R} \mid Q(T_n(G)) = \Omega(n^\beta) \},
q_M(T(G)) = \sup \{ \beta \in \mathbb{R} \mid Q_M(T_n(G)) = \Omega(n^\beta) \}.
\]
Proof. Suppose that \( R(T_n(G)) = O(n^\beta) \). Then,
\[
R(T_2(G)^\otimes N) = R(T_2^\otimes N(G)) = O(2^{N\beta}).
\]
so \( \omega(T(G)) \leq \beta \) by Theorem 1.1.16. Here we used \( T_2(G)^{\otimes N} \cong T_2^N(G) \). On the other hand, suppose that \( \omega(T(G)) < \beta \). Then,

\[
R(T_2^N(G)) = R(T_2(G)^{\otimes N}) = \mathcal{O}(2^{\beta N})
\]

by Theorem 1.1.16, so

\[
R(T_n(G)) \leq R(T_2^N(G)) = \mathcal{O}(2^{\beta N}) = \mathcal{O}(n^\beta)
\]

where \( N = \lceil \log_2 n \rceil \).

The proofs for \( q \) and \( q_M \) follow similarly from Theorem 1.1.16.

Note that \( \omega(T_r(G)) = \omega(T(G)) \log_2 r \).

**Definition 1.1.19.** Let \( \phi \in U_1 \otimes \cdots \otimes U_k \) be a \( k \)-tensor. A **flattening** of \( \phi \) is a 2-tensor (a matrix) obtained by any grouping of the tensor legs \( U_1, \ldots, U_k \) into two groups.

Flattenings are useful for obtaining bounds on tensor rank and related notions.

**Proposition 1.1.20.** Let \( \phi \) be a \( k \)-tensor and let \( A_\phi \) be any flattening of \( \phi \). Then \( R(A_\phi) \leq R(\phi) \) and \( Q(\phi) \leq Q(A_\phi) \). On the asymptotic level, we have the inequalities \( q(\phi) \leq q(A_\phi) = \log_2 Q(A_\phi) = \log_2 R(A_\phi) = \omega(A_\phi) \leq \omega(\phi) \).

**Proof.** A flattening of a simple \( k \)-tensor is a simple 2-tensor (a rank-1 matrix), and hence for an arbitrary \( k \)-tensor \( \phi \) we have \( R(A_\phi) \leq R(\phi) \). Similarly, \( Q(\phi) \leq Q(A_\phi) \).

For matrices, rank is multiplicative under tensor product and coincides with subrank. Therefore, for any \( n \in \mathbb{N} \),

\[
Q(A^{\otimes n}) \leq Q(A_\phi^{\otimes n}) = Q(A_\phi)^n = R(A_\phi)^n = R(A_\phi^{\otimes n}) \leq R(\phi^{\otimes n}).
\]

Taking the \( n \)th root, taking the logarithm \( \log_2 \) and letting \( n \) go to infinity, gives

\[
q(\phi) \leq q(A_\phi) = \log_2 Q(A_\phi) = \log_2 R(A_\phi) = \omega(A_\phi) \leq \omega(\phi),
\]

by Theorem 1.1.16, finishing the proof.

**Example 1.1.21.** A cut of a graph \( G = (V, E) \) is a partition of \( V \) into two disjoint nonempty sets. A max-cut is a cut with maximal number of edges crossing these two sets. A min-cut is a cut with minimal number of edges crossing these two sets. For any graph \( G \), let \( f(G) \) denote the size of a max-cut of \( G \), and \( g(G) \) the size of a min-cut. Flattening along the appropriate cuts yields

\[
Q(T(G)) \leq 2^{g(G)} \leq 2^{f(G)} \leq R(T(G))
\]
and thus
\[ q(T(G)) \leq g(G) \leq f(G) \leq \omega(T(G)). \tag{4} \]
While for a 2-tensor \( \phi \), we have \( Q(\phi) = R(\phi) \) and \( q(\phi) = \omega(\phi) \), the above observation gives us many examples of higher-order tensors for which this is not the case. For example, let \( \phi = T(C_k) \). Then \( g(C_k) = 2 \) and \( f(C_k) = k - 1 \).

Let us discuss the exponent in more detail.

**Definition 1.1.22.** Define the real number \( \omega := \omega(T(C_3)) \). This number is called the exponent of matrix multiplication.

In algebraic complexity theory, the number \( \omega \) is a measure for the asymptotic complexity of multiplying two \( n \times n \) matrices, and has been receiving much attention since the discovery of Strassen’s matrix multiplication algorithm [Str69]. See also the standard reference for algebraic complexity theory [BCS97]. The following bounds on \( \omega \) are the state of the art. The upper bound is by Le Gall [LG14] and the lower bound is by a standard flattening argument (Theorem 1.1.21).

**Theorem 1.1.23.** \( 2 \leq \omega \leq 2.3728639 \).

Computing \( \omega(T(G)) \) will be a hard task in general, since it includes computing the matrix multiplication exponent \( \omega \). One may however try to prove bounds on \( \omega(T(G)) \) in terms of the matrix multiplication exponent \( \omega \) or the dual exponent \( \alpha \). The number \( \alpha \) is defined as \( \sup \{ \gamma \in \mathbb{R} \mid \omega(1, 1, \gamma) = 2 \} \), where \( \omega(1, 1, \gamma) \) is defined as \( \inf \{ \beta \in \mathbb{R} \mid R(\langle n, n, \lfloor n^\gamma \rfloor \rangle) = O(n^\beta) \} \) and for \( n_1, n_2, n_3 \in \mathbb{N} \) the tensor \( \langle n_1, n_2, n_3 \rangle \) is the rectangular matrix multiplication tensor

\[
\langle n_1, n_2, n_3 \rangle := \sum_{i \in [n_1] \times [n_2] \times [n_3]} (b_{i_1} \otimes b_{i_2}) \otimes (b_{i_2} \otimes b_{i_3}) \otimes (b_{i_3} \otimes b_{i_1}) \\
\in (\mathbb{C}^{n_1} \otimes \mathbb{C}^{n_2}) \otimes (\mathbb{C}^{n_2} \otimes \mathbb{C}^{n_3}) \otimes (\mathbb{C}^{n_3} \otimes \mathbb{C}^{n_1}).
\]

The state of the art bounds on \( \alpha \) are \( 0.3029805 < \alpha \leq 1 \) [LG12]. In [CZ18] we used a technique called tensor surgery to obtain such a result for cycle graphs \( C_k \). Namely, let \( k \in \mathbb{N}_{\geq 1} \). Then

\[
\omega(T(C_k)) = k \quad \text{when } k \text{ is even,} \tag{5}
\]
\[
k - 1 \leq \omega(T(C_k)) \leq \frac{k - 1}{2} \omega \quad \text{when } k \text{ is odd.} \tag{6}
\]

Also, in terms of the dual exponent \( \alpha \), for odd \( k \),

\[
\omega(T(C_k)) \leq k - \alpha \left( 1 + \frac{1 - \alpha}{k - 1 + \alpha} \right) \leq k - \alpha.
\]

In particular, if \( \omega = 2 \) (equivalently \( \alpha = 1 \)), then \( \omega(T(C_k)) = k - 1 \) when \( k \) is odd. We will see a similar type of result for complete graphs in the next
subsection, see (9). We note that tensor surgery seems to work well for sparse
graphs like cycle graphs, and not so well for dense graphs like the complete
graph. However, one could use the complete graph tensors as starting tensors
in a tensor surgery procedure.

Definition 1.1.24. For any graph \( G = (V, E) \), we define the *exponent per edge*
\[ \tau(T(G)) := \omega(T(G))/|E|. \]
The letter \( \tau \) is borrowed from the Schönhage asymptotic sum inequality, also
known as the \( \tau \)-theorem.

For complete graphs, the exponent per edge has the following monotonicity
property.

Proposition 1.1.25. Let \( k \geq \ell \geq 2 \) be integers. Then
\[ \tau(T(K_k)) \leq \tau(T(K_\ell)). \]
Proof. Let \( \tau_\ell = \tau(T(K_\ell)) \). Label the vertices of \( K_k \) by \( 1, \ldots, k \). Then, for
any subgraph \( G \) in \( K_k \) isomorphic to \( K_\ell \) we have, with subscripts denoting
tensor leg positions,
\[ (T(G))_{V(G)} \otimes (1 \otimes \cdots \otimes 1)_{[k]\setminus V(G)}) \otimes^N \leq T^{\otimes(\ell^2 \tau_\ell N + o(N))} \]
and so
\[ \bigotimes_{G \subseteq K_k: G \cong K_\ell} (T(G))_{V(G)} \otimes (1 \otimes \cdots \otimes 1)_{[k]\setminus V(G)}) \otimes^N \leq T^{\otimes(\ell^2 \tau_\ell N + o(N)) (\frac{k}{\ell})}, \]
the tensor product taken over all subgraphs \( G \) in \( K_k \) isomorphic to \( K_\ell \). The
left-hand side is isomorphic to \( T(K_k)^{\otimes N(k-2)} \). Therefore, we have the upper
bound \( \omega(T(K_k)) \leq (\frac{k}{\ell}) (\frac{k-2}{\ell-2})^{-1} \tau_\ell = (\frac{k}{2}) \tau_\ell \), so \( \tau(T(K_k)) \leq \tau_\ell \). \( \square \)

Having discussed the exponent, we now wish to focus on the (monomial)
subexponent, and primarily on an important result about the (monomial)
subexponent of so-called tight 3-tensors.

Definition 1.1.26. Let \( \phi \) be an element of \( V_1 \otimes \cdots \otimes V_k \). Let \( B_1, \ldots, B_k \) be
bases for \( V_1, \ldots, V_k \) respectively. Write \( \phi \) in terms of these bases,
\[ \phi = \sum_{b_i \in B_i} \phi(b_1, \ldots, b_k) b_1 \otimes \cdots \otimes b_k \quad \text{with} \quad \phi(b_1, \ldots, b_k) \in \mathbb{C}. \]
Then the *support* of \( \phi \) with respect to \( B = (B_1, \ldots, B_k) \) is defined as the set of
\( k \)-tuples
\[ \text{supp}_B \phi := \{(b_1, \ldots, b_k) \in B_1 \times \cdots \times B_k \mid \phi(b_1, \ldots, b_k) \neq 0\}. \]
When the basis is clear from the context we will simply write \( \text{supp} \phi \). It is
often convenient to identify each \( B_i = \{b_1, b_2, \ldots, b_{|B_i|}\} \) with its index set
\( \{1, 2, \ldots, |B_i|\} \) by \( b_j \mapsto j \) so that \( \text{supp}_B \phi \) becomes a set of tuples of natural
numbers. (Sometimes the index set starts at 0 instead of 1.)
Then the marginal distribution of $P$ where

$$
\text{Theorem 1.1.31}
$$

These maps are injective and

Then we say $\phi$ is tight with respect to $\alpha = \alpha_1 \times \cdots \times \alpha_k$. We say $\phi$ is tight if it is tight with respect to $\alpha = \alpha_1 \times \cdots \times \alpha_k$ for some $\alpha : B_i \to \mathbb{Z}$.

The following lemma is easy to prove.

**Lemma 1.1.28.** Tensor products of tight tensors are tight in the tensor product basis. Therefore, any graph tensor $T(G)$ is tight in the standard basis.

**Example 1.1.29.** $T(C_k) = \sum_{i \in [2]^k} (b_{i_1} \otimes b_{i_2}) \otimes (b_{i_2} \otimes b_{i_3}) \otimes \cdots \otimes (b_k \otimes b_{i_1})$ is tight for any $k$ by Theorem 1.1.28. Explicitly, let

$$
\alpha_1 : b_{i_1} \otimes b_{i_2} \mapsto i_1 - 2(i_2 - 1),
\alpha_2 : b_{i_2} \otimes b_{i_3} \mapsto 2(i_2 - 1) - 4(i_3 - 1),
\vdots
\alpha_k : b_k \otimes b_{i_1} \mapsto 2^{k-1}(i_1 - 1) - i_1.
$$

These maps are injective and $\alpha_1(b_{i_1} \otimes b_{i_2}) + \cdots + \alpha_k(b_k \otimes b_{i_1}) = 0$ for any $i \in [2]^k$.

**Example 1.1.30.** For any partition $\lambda \vdash_d k$, the Dicke tensor $D_\lambda$ is a tight tensor. Namely, let $s = \sum_{j=1}^d \lambda_j \cdot j$, and let

$$
\alpha_1 : b_i \mapsto i,
\alpha_2 : b_i \mapsto i,
\vdots
\alpha_k : b_i \mapsto i - s.
$$

Then $\alpha_1(b_1) + \cdots + \alpha_k(b_k) = 0$ for any $(b_1, \ldots, b_k) \in \text{supp} D_\lambda$.

For tight 3-tensors, Strassen proved the following theorem. For a discrete probability distribution $P = (p_1, \ldots, p_n)$, let $H(P)$ be the Shannon entropy of $P$, which is defined by $H(P) := -\sum_{i=1}^n p_i \log_2 p_i$. Recall that if $P$ is a probability distribution on a set $X$ consisting of $k$-tuples $(x_1, \ldots, x_k)$, then the marginal distribution $P_i$ on the $i$th component $X_i = \{x_i \mid x \in X\}$ is defined by $P_i(y) := \sum_{x \in X : x_i = y} P(x)$.

**Theorem 1.1.31 ([Str91]).** Let $\phi$ be a 3-tensor which is tight in some basis $B$. Then

$$
q_M(\phi) = q(\phi) = \max_{P \in \mathcal{P}_\phi} \min \{H(P_1), H(P_2), H(P_3)\},
$$

where $\mathcal{P}_\phi$ consists of the probability distributions $P$ on $\text{supp}_B \phi$, and $P_i$ is the marginal distribution of $P$ on the $i$th component of $\text{supp}_B \phi$. 

12
Interestingly, Theorem 1.1.31 says that if $\phi$ is a tight tensor in basis $B$, then the subexponent $q(\phi)$ is a function of $\text{supp}_B \phi$, that is, $q(\phi)$ is independent of the coefficients of $\phi$ in basis $B$. This is not in general the case for higher-order tight tensors as we will see in Theorem 1.1.38.

**Example 1.1.32.** From Theorem 1.1.30 we know that $D_{(1,1,1)}$ is tight. Let $P$ be the uniform distribution on the support of $D_{(1,1,1)}$. Then each marginal $P_i$ is uniform on $\{b_1, b_2, b_3\}$. Therefore by Theorem 1.1.31, we have $q(D_{(1,1,1)}) \geq q_M(D_{(1,1,1)}) \geq \log_2 3$. On the other hand, $q(D_{(1,1,1)}) \leq \log_2 3$, because the tensor $D_{(1,1,1)}$ has a flattening of rank 3.

**Example 1.1.33.** From Theorem 1.1.28 and Theorem 1.1.29 we know that $\phi = T(C_3)$ is tight. Let $P$ be the uniform distribution on $\text{supp} \phi = \{(b_i \otimes b_j) (b_j \otimes b_k) (b_k \otimes b_i) | i,j,k \in [2]\}$. Then the marginals $P_i$ are uniform on $\{b_i \otimes b_j | i,j \in [2]\}$, and hence the Shannon entropy of each marginal is $H(P_i) = \log_2 4 = 2$. Therefore, by Theorem 1.1.31, $q(T(C_3)) \geq 2$. On the other hand, $q(T(C_3)) \leq 2$, because the tensor $T(C_3)$ has a flattening of rank 4.

There are two extensions of Theorem 1.1.31 to certain higher-order tensors, namely to W-state tensors (Theorem 1.1.34) and to (hyper)graphs with a certain connectedness property (Theorem 1.1.37), which includes cycle tensors. We begin with the extension to W-state tensors, obtained by generalising a construction of Coppersmith and Winograd and generalising Strassen’s support functionals. Define the binary entropy function $h(p)$ as $-p \log_2(p) - (1-p) \log_2(1-p)$ with $h(0) = h(1) = 0$.

**Theorem 1.1.34** (Vrana–Christandl [VC15]). Let $k \geq 2$. Then the subexponent of $D_{(1,k-1)}$ satisfies $q(D_{(1,k-1)}) = q_M(D_{(1,k-1)}) = h(1/k)$.

(The original proof contained a small mistake for which we provide a fix in Section 3.5.) It is conjectured in [VC15] that in general $q(D_\lambda) = q_M(D_\lambda) = H(\lambda/k)$, see also Theorem 1.3.3.

**Remark 1.1.35.** In [FK14] the Coppersmith–Winograd construction for $k = 3$ was used to prove lower bounds on the query complexity of testing triangle-freeness of Boolean functions. This construction was later extended to cover odd-cycle-freeness [HX15], thus independently proving part of the statement of Theorem 1.1.34.

**Example 1.1.36.** Let $\phi \in (C^2)^{\otimes 3}$ be nonzero. Then one of the following statements holds.

1. $\phi \cong T$ and $q(\phi) = 1$.
2. $\phi \cong D_{1,2}$ and $q(\phi) = h(1/3)$. 

13
3. \( \phi \cong T(s \leftrightarrow s), T(s \leftrightarrow \cdot \cdot), T(\cdot \leftrightarrow s) \) or \( T(\cdot \leftrightarrow \cdot \cdot) \) and \( q(\phi) = 0 \).

Indeed, it is well-known that any element \( \phi \in (\mathbb{C}^2)^{\otimes 3} \) is equivalent to precisely one of the six tensors listed above, see [DVC00] where \( \cong \) is called “equivalence under SLOCC”. In the terminology of the reference, the first tensor is in the GHZ-class, the second tensor is in the W-class and the remaining classes are called A-BC, AB-C, AC-B and A-B-C. In the first case, by definition \( q(\phi) = \omega(T, T)^{-1} = 1 \). In the second case, \( q(\phi) = q(D_{1,2}) = h(1/3) \) by Theorem 1.1.34. In the third case, every representative corresponds to a graph with min-cut of size 0, so \( q(\phi) \leq g(\phi) = 0 \) (see Theorem 1.1.21).

Recall that we defined \( g(G) \) to be the size of a minimum cut of \( G \) in Theorem 1.1.21. The following theorem is a special case of the result proved in [VC17].

**Theorem 1.1.37** (Vrana–Christandl [VC17]). Let \( G \) be a graph. Then

\[
q_{\text{TM}}(T(G)) = q(T(G)) = g(G).
\]

In particular, if \( k \geq 3 \), then the subexponent of \( T(C_k) \) satisfies \( q(T(C_k)) = 2 \).

**Example 1.1.38.** Strassen’s Theorem 1.1.31 does not generalize to \( k \geq 4 \) by simply replacing the right-hand side of (7) by \( \max_P \min \{ H(P_1), \ldots, H(P_k) \} \) as the following example shows. Let \( \phi \) be the 4-tensor

\[
\phi = T(s \leftrightarrow \cdot \cdot \cdot)
\]

\[
= (b_0 \otimes b_0 + b_1 \otimes b_1) \otimes (b_0 \otimes b_0 + b_1 \otimes b_1)
\]

\[
= b_0 \otimes b_0 \otimes b_0 \otimes b_0 + b_0 \otimes b_0 \otimes b_1 \otimes b_1
\]

\[
+ b_1 \otimes b_1 \otimes b_0 \otimes b_0 + b_1 \otimes b_1 \otimes b_1 \otimes b_1.
\]

This tensor is tight. Take for example

\[
\alpha_1 : b_i \mapsto i,
\]

\[
\alpha_2 : b_i \mapsto -i,
\]

\[
\alpha_3 : b_i \mapsto i,
\]

\[
\alpha_4 : b_i \mapsto -i.
\]

Let \( P \) be the uniform distribution on \( \text{supp} \phi \). Its marginals \( P_i \) are uniform on \( \{b_0, b_1\} \) so \( H(P_1) = 1 \). However, \( q(\phi) = 0 \), since \( \phi \) by construction has a flattening of rank 1.

This example also shows that \( q(\phi) \) cannot simply be a function of the support of \( \phi \), since, for \( 0 < p < 1 \), the tensor

\[
p(b_0 \otimes b_0 \otimes b_0 \otimes b_0 + b_0 \otimes b_0 \otimes b_1 \otimes b_1) + b_1 \otimes b_1 \otimes b_0 \otimes b_0 + b_1 \otimes b_1 \otimes b_1 \otimes b_1
\]

has no flattenings of rank at most 1, and hence \( q(\phi) \) is strictly positive (see Lemma 4 in [VC15]).

**Remark 1.1.39.** Theorem 1.1.38 suggests that to generalize Theorem 1.1.31 to \( k \)-tensors \( \phi \) one either has to find a stronger condition when \( k \geq 4 \) to guarantee that \( q(\phi) = \max_P \min \{ H(P_1), \ldots, H(P_k) \} \), or find a different lower (and possibly upper) bound on the subexponent \( q(\phi) \). Our Theorem 1.2.4 is a result of the second type.
1.2. Our results

This paper is motivated by the following problem on complete graph tensors. Recall that we defined the exponent per edge for the complete graph tensor as $\tau(T(K_k)) = \omega(T(K_k))/\binom{k}{2}$.

Problem 1.2.1. For $k \geq 4$, what is the value of $\tau(T(K_k))$?

First of all, it is not hard to prove the following bounds on $\tau(T(K_k))$.

For the complete graph $K_k$, the maximum cut size $f(K_k)$ is $k^2/4$ for even $k$ and $(k - 1)(k + 1)/4$ for odd $k$ (maximum cut is defined in Theorem 1.1.21). Then, flattening $T(K_k)$ along a max-cut yields a matrix of rank $2f(K_k)$ and therefore (Theorem 1.1.20)

$$\tau(T(K_k)) \geq \frac{f(K_k)}{\binom{k}{2}} = \begin{cases} 1/2 + 1/(2k) & \text{for odd } k \\ 1/2 + 1/(2(k - 1)) & \text{for even } k. \end{cases}$$

On the other hand, $\tau(T(K_3)) = \omega/3$ and thus by Theorem 1.1.25 we have $\tau(T(K_k)) \leq \omega/3$ for all $k \geq 3$. Plugging in the Le Gall upper bound $\omega \leq 2.3728639$ (Theorem 1.1.23), yields the “triangle covering” upper bound

$$\tau(T(K_k)) \leq 0.790955.$$  

(8)

(9)

Our aim is to improve on this upper bound.

Our main result is an upper bound on the exponent per edge of $T(K_k)$ that is independent of $\omega$.

Theorem 1.2.2. Let $k \geq 2$. For any $q \geq 1$, $\tau(T(K_k)) \leq \log_q \left( \frac{q + 2}{2^{qM(\phi)}} \right)$.

Our second result is a lower bound on the monomial subexponent $qM(\phi)$ of any tight tensor $\phi$.

Definition 1.2.3. Let $\phi$ be a $k$-tensor, $B$ a basis, $\Phi$ the corresponding support as in Theorem 1.1.26. Suppose $\phi$ is tight with respect to $\alpha$. For $R \subseteq \Phi \times \Phi$, define $r_\alpha(R)$ to be the rank of the matrix with rows

$$\{ \alpha(x) - \alpha(y) \mid (x, y) \in R \}.$$  

We will denote $r_\alpha(R)$ by $r(R)$ when the actual $\alpha$ is clear or not important. For any $i \in \{1, \ldots, k\}$, define

$$R_i := \{(x, y) \in \Phi \times \Phi \mid x_i = y_i\}.$$  

Theorem 1.2.4. Let $\phi$ be a $k$-tensor, $B$ a basis, $\Phi$ the corresponding support. Suppose $\phi$ is tight with respect to $\alpha$. Then,

$$qM(\phi) \geq \max_{P \subseteq \mathcal{P}} \left( H(P) - (k - 2) \max_{R \subseteq \mathcal{R}} \max_{Q \in \mathcal{Q}, (p_1, \ldots, p_k), R} \frac{H(Q) - H(P)}{r_\alpha(R)} \right),$$

where

$$\text{(10)}$$
• $\mathcal{P}_\Phi$ consists of all probability distributions $P$ on $\Phi$; and $P_1, \ldots, P_k$ are the marginal distributions of $P$ on the $k$ components respectively;

• $\mathcal{R}_\Phi$ consists of all subsets $R \subseteq \Phi \times \Phi$ that are not contained in the diagonal set $\Delta_\Phi := \{(x, x) \mid x \in \Phi\}$, and such that $R \subseteq \mathcal{R}_i$ for some $i \in \{1, \ldots, k\}$.

• $\mathcal{Q}_{\Phi, (P_1, \ldots, P_k), R}$ consists of all probability distributions $Q$ on $R$ whose marginals on the $2k$ components of $R$ satisfy $Q_i = Q_{k+i} = P_i$ for $i \in \{1, \ldots, k\}$.

(The symbols $\mathcal{P}$, $\mathcal{R}$, $\mathcal{Q}$ are script versions of the letters $P$, $R$ and $Q$.)

The lower bound in Theorem 1.2.4 extends the lower bound in Theorem 1.1.31. We will prove this in Section 3.5.

We will use Theorem 1.2.4 to compute the (monomial) subexponent of the weight-$(2, 2)$ Dicke tensor $D_{(2,2)}$. We give the proof in Section 3.5.

Corollary 1.2.5. $q_M(D_{(2,2)}) = q(D_{(2,2)}) = 1$.

Theorem 1.1.25, Theorem 1.2.2 with $k = 4$ and Theorem 1.2.5 together directly imply the following upper bound on the exponent per edge of the complete graph tensor $T(K_k)$ for any $k \geq 4$.

Corollary 1.2.6. For $k \geq 4$, $\tau(T(K_k)) \leq \min_{q \geq 2} \log_q \left(\frac{q + 2}{2}\right) = \log_7(9/2)$ which is approximately 0.772943.

Note that 0.772943 is strictly smaller than the triangle cover upper bound 0.790955 from (9).

In the following table we summarize, for small $k$: the flattening lower bound on $\omega(T(K_k))$; the upper bound from Theorem 1.2.6 on $\omega(T(K_k))$ (and Le Gall’s upper bound for $k = 3$); the trivial upper bound on $\omega(T(K_k))$ given by the number of edges; and the resulting bounds on $\tau(T(K_k))$.

<table>
<thead>
<tr>
<th>$k$</th>
<th>$\omega(T(K_k))$</th>
<th>$\binom{k}{2}$</th>
<th>$\tau(T(K_k))$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>lower</td>
<td>upper</td>
<td>lower</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2.37287</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4.63766</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>7.72943</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>11.5942</td>
<td>15</td>
</tr>
<tr>
<td>7</td>
<td>12</td>
<td>16.2319</td>
<td>21</td>
</tr>
<tr>
<td>8</td>
<td>16</td>
<td>21.6425</td>
<td>28</td>
</tr>
<tr>
<td>9</td>
<td>20</td>
<td>27.8260</td>
<td>36</td>
</tr>
<tr>
<td>10</td>
<td>25</td>
<td>34.7825</td>
<td>45</td>
</tr>
</tbody>
</table>

16
1.3. Questions

We discuss three open questions related to our results. Our first question is about complete graph tensors. From (8) we know that
\[
\tau(T(K_{k+1})) \geq \tau(T(K_k)) \geq \frac{(k-1)(k+1)}{4(k/2)}
\]
holds for all odd \(k\). The lower bound goes to \(1/2\) when \(k\) goes to infinity. On the other hand, if \(\omega = 2\), then \(\tau(T(K_k)) \leq \frac{2}{3}\) holds for all \(k \geq 3\) (Theorem 1.1.25), and in particular \(\tau(T(K_3)) = \tau(T(K_4)) = \frac{2}{3}\). We thus ask the following question.

**Question 1.3.1.** Is there a \(k \geq 5\) such that \(\tau(T(K_k)) < \frac{2}{3}\)?

Our second question is a more precise version of the above question for general graph tensors.

**Question 1.3.2.** Is it true that for every graph \(G\), we have \(\omega(T(G)) = f(G)\)?

The answer is known to be yes only for bipartite graphs. (This follows directly from the lower bound in (4).) If the answer to Theorem 1.3.2 is yes, then the answer to Theorem 1.3.1 is yes. For \(G = C_3\), this question specializes to the long-standing open question whether the exponent of matrix multiplication \(\omega\) equals 2. It might therefore be interesting to ask Theorem 1.3.2 with the additional assumption that \(\omega = 2\). The answer is then known to be yes for all odd cycles (see (6)), the bipartite graphs, the complete graph \(K_4\) and graphs that are composed of these in a certain (natural) way; namely by taking the disjoint union and then identifying pairs of nonadjacent vertices of choice.

Our third question is about the (monomial) subexponent of Dicke tensors. In [VC17] the upper bound \(q(D_\lambda) \leq H(\lambda/k)\) was proven for any \(\lambda \vdash k\), and the following problem was posed.

**Question 1.3.3.** Is it true that for every partition \(\lambda \vdash k\),
\[
q_M(D_\lambda) = q(D_\lambda) = H(\lambda/k)
\]
(11)

For \(k = 3\) this question was affirmatively answered in [Str91] (see Theorem 1.1.31 above), and for \(\lambda = (k - 1, 1)\) and any \(k \geq 3\) this question was affirmatively answered in [VC17] (see Theorem 1.1.34 above). Theorem 1.2.4 extends both results and covers more cases, including \(\lambda = (2, 2)\), see Theorem 1.2.5. We also numerically confirmed (11) for \(\lambda = (\ell, \ell)\) and \(2 \leq \ell \leq 1000\) using Theorem 1.2.4. We conjecture that Theorem 1.2.4 is strong enough to prove the lower bound \(q_M(D_\lambda) \geq H(\lambda/k)\) for any \(\lambda \vdash k\).
1.4. Outline

The rest of this paper is organized as follows. In Section 2 we will prove Theorem 1.2.2 on the exponent of $T(K_k)$. In Section 3 we will prove Theorem 1.2.4 on the monomial subexponent of tight tensors and we will compute the monomial subexponent of the weight-$(2, 2)$ Dicke tensor.

2. Upper bound on the exponent of the complete graph tensor

The main structure of the proof of Theorem 1.2.2 is a generalization of a construction of Strassen [Str86] which was improved by Coppersmith and Winograd [CW87], which involves finding a “cheap” starting tensor, generalizing Schönhage’s asymptotic sum inequality [Sch81] and choosing a good block decomposition.

2.1. Preliminaries

We start by discussing border rank in more detail, then we introduce the generalized asymptotic sum inequality and finally we review block decompositions of tensors.

Let $\phi \in V_1 \otimes \cdots \otimes V_k$ be a tensor.

**Definition 2.1.1.** Let $R_h(\phi)$ be the smallest number $r$ such that there are matrices $A_1(\varepsilon) : \mathbb{C}^r \to V_1$, ..., $A_k(\varepsilon) : \mathbb{C}^r \to V_k$ with entries in $\mathbb{C}[\varepsilon]$, such that $(A_1(\varepsilon) \otimes \cdots \otimes A_k(\varepsilon)) T_r = \varepsilon^h \phi + O(\varepsilon^{h+1})$. Here, $O(\varepsilon^{h+1})$ denotes any expression of the form $\sum_{i=1}^{\ell} \varepsilon^{h+i} \phi_i$ for some $\ell$.

**Theorem 2.1.2** ([BCS97, Theorem 20.24]). $R(\phi) = \min_h R_h(\phi)$.

It is well-known that rank and border rank are related as follows.

**Proposition 2.1.3.** Let $k, h \in \mathbb{N}$. Let $c_h = \binom{h+k-1}{k-1}$. For all $m_1, \ldots, m_k \in \mathbb{N}$ and all tensors $\phi \in \mathbb{C}^{m_1} \otimes \cdots \otimes \mathbb{C}^{m_k}$, we have $R(\phi) \leq c_h R_h(\phi)$. Note that for fixed $k$, the number $c_h$ is upper bounded by a polynomial in $h$; $c_h \leq (h+1)^{k-1}$.

**Proof.** Let $\phi$ be a tensor in $\mathbb{C}^{m_1} \otimes \cdots \otimes \mathbb{C}^{m_k}$ with $R_h(\phi) = r$. Then there are vectors $v_i^j \in (\mathbb{C}[\varepsilon])^{m_j}$ such that

$$\sum_{i=1}^{r} v_1^i \otimes \cdots \otimes v_k^i = \varepsilon^h \phi + O(\varepsilon^{h+1}).$$

Without loss of generality the highest power of $\varepsilon$ in each $v_i^j$ is $h$. Decomposing every $v_i^j$ into $\varepsilon$-homogeneous components $v_i^j = \sum_{a_j=0}^{h} \varepsilon^{a_j} v_i^j(a_j)$, and collecting powers of $\varepsilon$ gives

$$\sum_{i=1}^{r} \sum_{a \in [h]^k} \varepsilon^{a_1 + \cdots + a_k} v_i^1(a_1) \otimes \cdots \otimes v_i^k(a_k) = \varepsilon^h \phi + O(\varepsilon^{h+1}).$$
Taking only the summands such that \( a_1 + \cdots + a_k = h \) gives a rank decomposition of \( \phi \). There are \( \binom{h+k-1}{k-1} \) such summands. Therefore, the statement of the proposition holds for \( c_h = \binom{h+k-1}{k-1} \) which is at most \((h+1)^{k-1}\). \(\square\)

As a consequence of Theorem 2.1.3 we can upper bound the exponent of a tensor by the border rank of that tensor. (The following proposition also follows from the more general Theorem 1.1.8.)

**Proposition 2.1.4.** Let \( \psi \) be a tensor. Then

\[
\omega(\psi) \leq \log_2 R(\psi).
\]

**Proof.** Suppose \( R(\psi) = r \). Let \( h \in \mathbb{N} \) such that \( R_h(\psi) = r \) (Theorem 2.1.2). Then for all \( s \in \mathbb{N}, R_{hs}(\psi\otimes^s) \leq r^s \). Therefore, \( R(\psi\otimes^s) \leq c_h r^s \) (Theorem 2.1.3). Thus the exponent \( \omega(\psi) \) is at most \( \frac{1}{s} \log_2 r^s + \frac{1}{s} \log_2 c_h \), which converges to \( \log_2 r \) when \( s \) goes to infinity. \(\square\)

Let \( G = (V,E) \) be a graph and let \( f : E \to \mathbb{N} \) be a function that assigns to every edge a natural number. We define a “nonuniform” version of \( T(G) \) as follows. Define \( T_f(G) \) by

\[
T_f(G) := \bigotimes_{e \in E} \sum_{i \in [f(e)]} (b_i \otimes \cdots \otimes b_i)_e \otimes (1 \otimes \cdots \otimes 1)_{V\setminus e}.
\]

Here, the large tensor product inputs and outputs \( |V| \)-tensors. In algebraic complexity theory, the tensor \( T_f(C_3) \) with \( f(1) = n_1, f(2) = n_2, f(3) = n_3 \) is denoted by \( \langle n_1, n_2, n_3 \rangle \). This tensor corresponds to the bilinear map that multiplies an \( n_1 \times n_2 \) matrix with an \( n_2 \times n_3 \) matrix. We view the set \( \{f : E \to \mathbb{N}\} \) as a group under pointwise multiplication, so that we can write

\[
T_f(G) \otimes T_g(G) \cong T_{fg}(G). \quad (12)
\]

Equation \((12)\) generalises the self-reducibility property of matrix multiplication tensors: \( \langle m_1, m_2, m_3 \rangle \otimes \langle m_1, m_2, m_3 \rangle \cong \langle m_1 n_1, m_2 n_2, m_3 n_3 \rangle \).

Recall that an **automorphism** of a graph \( G = (V,E) \) is a permutation \( \sigma \) of \( V \) such that for all \( u, v \in V \) the pair \( (u,v) \) is in \( E \) if and only if \( \sigma \cdot (u,v) := (\sigma(u), \sigma(v)) \) is in \( E \). The automorphisms form a group \( \Gamma \) under composition. The group \( \Gamma \) thus acts on \( V \) and on \( E \). We say \( G \) is **edge-transitive** if the action of \( \Gamma \) on \( E \) is transitive, meaning that for any two edges \( e_1, e_2 \in E \) there is a permutation \( \sigma \in \Gamma \) such that \( \sigma \cdot e_1 = e_2 \).

**Example 2.1.5.** The automorphism group of the cycle graph \( C_k \) is the dihedral group with \( 2n \) elements. The automorphism group of the complete graph \( K_k \) is the symmetric group \( S_k \). Both graphs are edge-transitive.

19
Theorem 2.1.6 (Generalized asymptotic sum inequality). Let $G = (V,E)$ be an edge-transitive graph. Suppose $r > p$. Suppose $\phi_1, \ldots, \phi_p$ are tensors from $\{ T_f(G) \mid f : E \to \mathbb{N}; \prod_{e \in E} f(e) \geq 2 \}$ such that

$$R(\phi_1 \oplus \cdots \oplus \phi_p) \leq r.$$  

Define $\tau$ by $\sum_{i=1}^p (\prod_{e \in E} f_i(e))^r = r$. Then $\tau(T(G)) \leq \tau$.  

In particular, if $\phi_1, \ldots, \phi_p \in \{ T_f(G) \mid f : E \to \mathbb{N}; \prod_{e \in E} f(e) = q \}$ for some integer $q \geq 2$, then we have $\tau(T(G)) \leq \log_q (R(\phi_1 \oplus \cdots \oplus \phi_p)/p)$.

Our proof of Theorem 2.1.6 follows the structure of the proof for the cycle graph case in [BCZ17] which builds upon the exposition in [Blä13].

Proposition 2.1.7. Let $G = (V,E)$ be an edge-transitive graph. Let $f : E \to \mathbb{N}$ be a function and $N = \prod_{e \in E} f(e)$. Assume that $N \geq 2$. Then,

$$\tau(T(G)) \leq \log N R(T_f(G)).$$

Proof. Let $\Gamma$ be the automorphism group of $G$. For any edge $e \in E$ define the stabilizer subgroup $\Gamma_e := \{ \sigma \in \Gamma \mid \sigma \cdot e = e \}$. Since $G$ is edge-transitive all subgroups $\Gamma_e$ are conjugates and thus have the same cardinality, say $C$. This implies that, for any $e_1, e_2 \in E$, the set $\{ \sigma \in \Gamma \mid \sigma \cdot e_1 = e_2 \}$ has cardinality $C$. By the orbit-stabilizer theorem the number $C$ satisfies $|\Gamma| = C|E|$. The group $\Gamma$ acts naturally on $T_f(G)$ by permuting tensor legs. The tensor product of all elements $\sigma \cdot T_f(G)$ for $\sigma \in \Gamma$ equals $T_{NC}(G)$. We thus have, using Theorem 2.1.4,

$$\omega(T(G)) = \frac{\omega(T_{NC}(G))}{C \log_2 N} \leq \frac{|\Gamma| \log_2 R(T_f(G))}{C \log_2 N} = \frac{|E| \log_2 R(T_f(G))}{\log_2 N},$$

finishing the proof. In the above we used $\omega(T_r(G)) = \omega(T(G)) \log_2 r$. 

We need the following upper bound on the rank of the $a$-fold direct sum of the $s$th tensor power of a tensor.

Lemma 2.1.8. Let $a, b \in \mathbb{N}_{\geq 1}$. Let $\phi$ be a tensor such that $R(\phi^{\otimes a}) \leq b$. Then for all $s \in \mathbb{N}$, $R((\phi^{\otimes s})^{\otimes a}) \leq [b/a]^s a$.

Proof. We prove the lemma by induction over $s$. The base case $s = 1$ follows from the assumption. For the induction step, we have

$$(\phi^{\otimes (s+1)})^{\otimes a} = \phi^{\otimes s} \otimes \phi^{\otimes a} \leq \phi^{\otimes s} \otimes T_b = (\phi^{\otimes s})^{\otimes b},$$
and thus, by the induction hypothesis,
\[
R((φ^{(s+1)}a)⊗b) ≤ R((φ^s)⊗b)
≤ R((φ^s)⊗(a[b/a])) = R(((φ^s)a)⊗[b/a])
≤ [b/a]s[a] = [b/a]s+1a,
\]
proving the lemma.

Theorem 2.1.8 can equivalently be phrased as follows. Let \(φ\) be a tensor such that \(φ⊗T_a ≤ T_b\). Then for all \(s ∈ \mathbb{N}\) we have \(φ^s⊗T_a ≤ (T_b/a)^s⊗T_a\).

Formulated in the language of Strassen’s semiring of \(k\)-tensors, Theorem 2.1.8 says:
\(a · φ ≤ b ⇒ ∀sa · φ ≤ [b/a]^s · a\).

Theorem 2.1.7 generalizes to the following inequality relating upper bounds on the exponent of a direct power \(T_f(G) ⊕ a\) to upper bounds on the exponent of \(T(G)\).

**Proposition 2.1.9.** Let \(G = (V, E)\) be an edge-transitive graph. Let \(f : E → \mathbb{N}\) be a function and \(N = \prod_{e ∈ E} f(e)\). Assume that \(N ≥ 2\). Then, \(τ(T(G)) ≤ \log N [R(T_f(G)^{⊗a})/a]\), for any \(a ≥ 1\).

**Proof.** Let \(b = R((T_f(G))^{⊗a})\). Then by Theorem 2.1.8 we have the inequality \(R((T_f(G))^{⊗a}) ≤ [b/a]^s · a\), where \(f^s\) denotes taking the pointwise \(s\)th power. Therefore, by Theorem 2.1.7,
\[
\begin{align*}
τ(T(G)) &≤ \log N, R(T_f(G)) \\
&≤ \log N, R((T_f(G)^{⊗a})) \\
&≤ \log N, [b/a]^s \\
&= \frac{s \log [b/a] + \log(a)}{s \log(N)},
\end{align*}
\]
which goes to \(\log N [b/a]\) when \(s\) goes to infinity.

**Proof of Theorem 2.1.6.** Suppose \(r = R(\bigoplus_{i=1}^p T_{f_i}(G))\). This implies that there is an \(h ∈ \mathbb{N}\) such that \(R_h(\bigoplus_{i=1}^p T_{f_i}(G)) = r\). Taking the \(s\)th power gives \(R_{hs}(\bigoplus_{i=1}^p T_{f_i}(G)^{⊗s}) ≤ r^s\). We expand the tensor power to get
\[
R_{hs}(\bigoplus_{σ} (T_{f_{σ1}}^1(G) ⊗ \cdots ⊗ T_{f_{σp}}^p(G)))^{⊗(s)} ≤ r^s,
\]
where the first direct sum is over all \(p\)-tuples \(σ = (σ_1, \ldots, σ_p)\) of nonnegative integers with sum \(s\) (this is by the multinomial theorem). We can also write this inequality as
\[
R_{hs}(\bigoplus_{σ} (T_{f_{σ_1}^1}^1 \cdots T_{f_{σ_p}^p}(G)))^{⊗(s)} ≤ r^s.
\]
By Theorem 2.1.3 there exists a number $c_{hs}$ which is at most a polynomial in $h$ and $s$ such that

$$R\left(\bigoplus_{\sigma} (T_{f_{1}^{\sigma_{1}} \ldots f_{p}^{\sigma_{p}}(G))^{\sqcup}(\sigma))\right) \leq c_{hs} r^{s}. \quad (13)$$

Define $\tau$ by $\sum_{i=1}^{p} (\prod_{e \in E} f_{i}(e))^{\sigma} = r$. Then

$$\sum_{\sigma} \binom{s}{\sigma} \left(\prod_{e \in E} f_{i}(e)^{\sigma_{i}}\right)^{\tau} = r^{s}, \quad (14)$$

by the multinomial theorem again. In this sum, consider the maximum summand and fix $\sigma$ to be the corresponding $\sigma$ for the remainder of the proof.

Define $f : E \to \mathbb{N}$ by $f(e) := \prod_{i} f_{i}(e)^{\sigma_{i}}$. Let $a := \binom{s}{\sigma}$ and let $b := c_{hs} r^{s}$.

Equation (13) implies $R(T_{f}(G) \oplus a) \leq b$ which by Theorem 2.1.9 implies

$$\tau(T(G)) \leq \log_{N} \left(\frac{R(T_{f}(G) \oplus a)}{a}\right) \leq \log_{N} \left(\frac{b}{a}\right), \quad (15)$$

where $N := \prod_{e} f(e)$. There are $(s+p-1)^{(p-1)} \leq (s+1)^{p-1}$ $p$-tuples $\sigma$ with sum $s$, so there are that many summands in (14). We thus lower bound the maximum summand by the average of the summands as follows,

$$aN^{\tau} \geq \frac{r^{s}}{(s+p-1)^{(p-1)}} \geq \frac{r^{s}}{(s+1)^{p-1}}, \quad (16)$$

Manipulating (16) gives

$$\left\lfloor \frac{b}{a} \right\rfloor \leq \frac{c_{hs} r^{s}}{a} + 1 \leq N^{\tau} (s+1)^{p-2}2c_{hs}$$

which we plug into (15) to get

$$\tau(T(G)) \leq \frac{\tau \log_{2} N + (p-1) \log_{2}(s+1) + \log_{2}(2c_{hs})}{\log_{2} N},$$

$$\leq \tau + \frac{(p-1) \log_{2}(s+1) + \log_{2}(2c_{hs})}{\log_{2} N},$$

which goes to $\tau$ when $s$ goes to infinity, because (16) implies

$$N^{\tau} \geq \frac{r^{s}}{(s+1)^{p-1}a} \geq \frac{r^{s}}{(s+1)^{p-1}p^{s}}$$

and therefore $\log_{2} N$ is at least $s(\log_{2} r - \log_{2} p) - (p-1) \log_{2}(s+1)$, and we assumed $r > p$. \(\Box\)

We are now ready to discuss the final ingredient for the proof of Theorem 1.2.2, which is a generalization of block decompositions of matrices to block decompositions of tensors.
Definition 2.1.10 (Set partition and tensor partition). Let $A$ be a finite set. We define a partition of $A$ as a collection $\mathcal{A} = \{A^i\}$ of disjoint subsets of $A$ whose union is $A$. Let $B_1, \ldots, B_k$ be finite sets, and let $\mathcal{B}_1, \ldots, \mathcal{B}_k$ be partitions of each set respectively. Then we say $\mathcal{B} = \mathcal{B}_1 \times \cdots \times \mathcal{B}_k$ is a product partition of $B_1 \times \cdots \times B_k$.

Let $\phi \in V_1 \otimes \cdots \otimes V_k$ be a tensor and $B_1, \ldots, B_k$ orthogonal bases for $V_1, \ldots, V_k$ respectively. Let $\mathcal{B}$ be a product partition of $B_1 \times \cdots \times B_k$. Let $I \in \mathcal{B}$. Then we define $\phi_I$ as the orthogonal projection of $\phi$ onto the linear space spanned by $I$. These smaller tensors $\phi_I$ with $I \in \mathcal{B}$ we think of as making up the inner structure of $\phi$ with respect to $\mathcal{B}$. We define the outer structure of $\phi$ with respect to $\mathcal{B}$ to be the tensor $\phi_{\mathcal{B}}$ with entries indexed by $I \in \mathcal{B}$ such that $\phi_{\mathcal{B}}$ has a 1 at position $I$ if $\phi_I$ is not the zero tensor, and a 0 otherwise.

Definition 2.1.11. Let $\mathcal{B}$ be a product partition of $B_1 \times \cdots \times B_k$ and let $\mathcal{C}$ be a product partition of $C_1 \times \cdots \times C_k$. Define the product partition $\mathcal{B} \otimes \mathcal{C}$ of $(B_1 \times C_1) \times \cdots \times (B_k \times C_k)$ by

$$\mathcal{B} \otimes \mathcal{C} := \{I \otimes J | I \in \mathcal{B}, J \in \mathcal{C}\},$$

where $I \otimes J = \{(b_1, c_1), \ldots, (b_k, c_k) | b \in I, c \in J\}$.

The following proposition follows directly from the definition.

Proposition 2.1.12. Let $\phi \in U_1 \otimes \cdots \otimes U_k$ and $\psi \in V_1 \otimes \cdots \otimes V_k$. Let $B_i$ be a basis of $U_i$ and let $C_i$ be basis of $V_i$. Let $\mathcal{B}$ be a product partition of $B_1 \times \cdots \times B_k$ and let $\mathcal{C}$ be a product partition of $C_1 \times \cdots \times C_k$. Then,

$$(\phi \otimes \psi)_{\mathcal{B} \otimes \mathcal{C}} = \phi_{\mathcal{B}} \otimes \psi_{\mathcal{C}},$$

and $(\phi \otimes \psi)_{I \otimes J} = \phi_I \otimes \psi_J$ for all $I \in \mathcal{B}$, $J \in \mathcal{C}$.

2.2. Proof of Theorem 1.2.2

We define a modification of the “Coppersmith–Winograd tensor”, which is a higher-dimensional version of the weight-$(2, k - 2)$ Dicke tensor.

Definition 2.2.1. Let $k, q \geq 2$ be integers. Let $b_0, b_1, \ldots, b_q$ be the standard basis of $\mathbb{C}^{q+1}$. Let $\text{CW}_q^k$ be the $k$-tensor

$$\text{CW}_q^k := \sum_{i \in \{0,1,\ldots,q\}^k} b_{i_1} \otimes \cdots \otimes b_{i_k} \in (\mathbb{C}^{q+1})^\otimes k$$

where the sum goes over all $i$ that contain exactly two identical non-zero entries. One can also think of this tensor as a sum of tensors $T_q(e)$ with $e$
going over the edges of the complete graph $K_k$, that is, using subscripts to
denote the position of tensor legs,
\[
\text{CW}_q^k = \sum_{e \in E(K_k)} \sum_{i=1}^q (b_i \otimes b_i)_e \otimes (b_0 \otimes b_0 \otimes \cdots \otimes b_0)_{[k]}_e \in (\mathbb{C}^{q+1})^{\otimes k}. \tag{17}
\]

Note that in (17) the sums runs over $i$ from 1 to $q$.

The $\text{CW}_q^k$-tensor should be thought of as a cheap tensor, in the following
border rank sense.

\textbf{Lemma 2.2.2.} The border rank of $\text{CW}_q^k$ is at most $q + 2$.

\textit{Proof.} We have
\[
\sum_{i=1}^q (b_0 + \varepsilon^2 b_i)^{\otimes k} - \left( b_0 + \varepsilon^3 \sum_{i=1}^q b_i \right)^{\otimes k} + (1 - q\varepsilon) b_0^{\otimes k} = \varepsilon^5 \text{CW}_q^k + O(\varepsilon^6).
\]

Therefore, $\mathcal{R}(\text{CW}_q^k) \leq q + 2$. \hfill \Box

\textbf{Theorem 1.2.2} (Repeated). Let $k \geq 2$. For any $q \geq 1$,
\[
\tau(T(K_k)) \leq \log_q \left( \frac{q + 2}{2^{qM(D(k-2,2))}} \right).
\]

\textbf{Proof of Theorem 1.2.2}. Define a product partition $\mathcal{B} = B_1 \times \cdots \times B_k$ of
the standard basis of $(\mathbb{C}^{q+1})^{\otimes k}$ by $B_j := \{ \{ b_0 \}, \{ b_1, \ldots, b_q \} \}$ for all $j$. Then
the outer structure $(\text{CW}_q^k)_{\mathcal{B}}$ equals
\[
(\text{CW}_q^k)_{\mathcal{B}} = \sum_{e \in E(K_k)} (b_1 \otimes b_1)_e \otimes (b_0 \otimes b_0 \otimes \cdots \otimes b_0)_{[k]}_e
\]
which is the Dicke tensor $D_{(k-2,2)}$. (We implicitly reorder $(k-2,2)$ to $(2,k-2)$
when $k \leq 3$.) The inner structure of $\text{CW}_q^k$ with respect to $\mathcal{B}$ consists of tensors
\[
\sum_{i=1}^q (b_i \otimes b_i)_e \otimes (b_0 \otimes b_0 \otimes \cdots \otimes b_0)_{[k]}_e \quad \text{with} \quad e \in E(K_k).
\]

By Theorem 1.1.17, we have the following monomial degeneration on the outer structure level:
\[
T^{\otimes (qM(D_{(k-2,2)})^{s-o(s)})} \preceq_M ((\text{CW}_q^k)^{\otimes s})_{\mathcal{B}^{\otimes s}}. \tag{18}
\]

Since the degeneration in (18) is monomial, each nonzero entry in the lowest-
degree part of the tensor on the left-hand side corresponds to an inner
structure tensor of $(\text{CW}_q^k)^{\otimes s}$ with respect to $\mathcal{B}^{\otimes s}$. Each such inner structure
tensor is a product of \( s \) inner structure tensors of \( CW_q^k \) with respect to \( B \) (Theorem 2.1.12). Therefore,

\[
\phi_1 \oplus \cdots \oplus \phi_p \leq (CW_q^k)^{\otimes s}
\]

with \( p = 2^{m_2(D_{(k-2,2)})s-o(s)} \) and \( \phi_i \in \{ T_f(K_k) \mid \prod_{e \in E} f(e) = q^3 \} \). Then, by Theorem 2.2.2 we have

\[
R(\phi_1 \oplus \cdots \oplus \phi_p) \leq R((CW_q^k)^{\otimes s}) \leq (q+2)^s
\]

We apply the “in particular” case of the generalized asymptotic sum inequality Theorem 2.1.6 with \( r \) defined as \((q+2)^s\) to the graph \( G = K_k \) to obtain

\[
\tau(T(K_k)) \leq \log_q \frac{(q+2)^s}{2^{qM(D_{(k-2,2)})s-o(s)}}
\]

\[
\leq \log_q \frac{q+2}{2^{qM(D_{(k-2,2)})-o(1)}}.
\]

Letting \( s \) go to infinity yields the required inequality. \( \square \)

**Summary.** We wish to give a short summary of the generalized Coppersmith–Winograd method as employed above, for \( k = 4 \). Let \( K_4 = (E,V) \) be the tetrahedron. We aim to get a good upper bound on the tensor rank of a tensor power of the goal tensor:

\[
\text{goal tensor} = T \left( \begin{array}{c}
\otimes
\end{array} \right)
\]

\[
= \bigotimes_{e \in E} \sum_{i \in [2]} (b_i \otimes b_i)_e \otimes (b_0 \otimes b_0)_{V \setminus \{e\}}.
\]

To this end, we pick the following starting tensor, for which we have a good border rank upper bound (Theorem 2.2.2):

\[
\text{starting tensor} = T_q \left( \begin{array}{c}
\otimes
\end{array} \right) + T_q \left( \begin{array}{c}
\otimes
\end{array} \right) + T_q \left( \begin{array}{c}
\otimes
\end{array} \right)
\]

\[
+ T_q \left( \begin{array}{c}
\otimes
\end{array} \right) + T_q \left( \begin{array}{c}
\otimes
\end{array} \right) + T_q \left( \begin{array}{c}
\otimes
\end{array} \right)
\]

\[
= \sum_{e \in E} \sum_{i \in [q]} (b_i \otimes b_i)_e \otimes (b_0 \otimes b_0)_{V \setminus \{e\}}.
\]
We now choose a product partition of the standard basis for the space that the starting tensor lives in. This partitions the starting tensors into blocks (like we can partition matrices into blocks). The partition we use is \( \{\{b_0, \ldots, b_q\}\}^4 \). The partitioned tensor has, besides the zero block, the following blocks:

\[
\text{inner structure} = \left\{ T_{q}^{(1)}(\begin{array}{c} 1 \\ \end{array}), T_{q}^{(2)}(\begin{array}{c} 2 \\ \end{array}), T_{q}^{(3)}(\begin{array}{c} 3 \\ \end{array}) \right. \\
\left. T_{q}^{(4)}(\begin{array}{c} 4 \\ \end{array}), T_{q}^{(5)}(\begin{array}{c} 5 \\ \end{array}), T_{q}^{(6)}(\begin{array}{c} 6 \\ \end{array}) \right\} \\
= \left\{ \sum_{i \in [q]} (b_i \otimes b_i)e \otimes (b_0 \otimes b_0)_{V\setminus\{e\}} \mid e \in E \right\}.
\]

The following outer structure tensor tells us how the nonzero blocks are positioned, in block coordinates:

\[
\text{outer structure} = D_{(2,2)} \\
= \sum_{e \in E} (b_1 \otimes b_1)e \otimes (b_0 \otimes b_0)_{V\setminus\{e\}}.
\]

We can degenerate \( D_{(2,2)}^{\otimes s} \) to \( \sum_{i \in [p]} b_i \otimes b_i \otimes b_i \otimes b_i \) for \( p = 2^{\text{RM}(D_{(2,2)})s-o(s)} \) by a monomial degeneration (Theorem 1.1.15). This implies that we can degenerate the \( s \)th power of the starting tensor to a direct sum of nonuniform goal tensors, but such that for each nonuniform goal tensor the product of the edge weights equals \( q^s \) (by Theorem 2.1.12):

\[
T_{f_1}(\begin{array}{c} 1 \\ \end{array}) \oplus T_{f_2}(\begin{array}{c} 2 \\ \end{array}) \oplus \cdots.
\]

We then apply the generalized asymptotic sum inequality Theorem 2.1.6, whose construction consists of the following two steps. We first restrict a power of the above tensor to a direct power of a single nonuniform goal tensor

\[
T_f(\begin{array}{c} 1 \\ \end{array})^{\oplus a}.
\]

Second, another tensor power and symmetrization procedure gives a power of the goal tensor:

\[
T(\begin{array}{c} 1 \\ \end{array})^{\otimes N}.
\]

3. Lower bound on the monomial subexponent of tight tensors

We first discuss some results on \( k \)-average-free sets, linear combinations of independent uniformly random variables and types of sequences, that we will use in the proof of Theorem 1.2.4. Then we describe a basic procedure to restrict any tensor to a tensor of the form \( T_r \) with a monomial restriction. Then, we give the proof for Theorem 1.2.4. Next we discuss some computational aspects of Theorem 1.2.4. Finally, we give some applications.
3.1. Preliminaries

We begin with a result on $k$-average-free sets which is essentially due to Salem and Spencer [SS42]. For a self-contained proof, see [VC15, Lemma 10].

**Definition 3.1.1.** A subset $A \subseteq \mathbb{N}$ is $k$-average-free if $a_1, \ldots, a_k, y \in A$ and $a_1 + \cdots + a_k = ky$ implies $a_1 = \cdots = a_k = y$. For $N \in \mathbb{N}$, let $\nu_k(N)$ be the maximum size of a $k$-average-free set in $[N] = \{1, 2, \ldots, N\}$.

**Proposition 3.1.2.** For any fixed $k$, we have $\nu_k(N) = N^{1-o(1)}$ as $N \to \infty$.

The following statement about linear combinations of independent uniformly random variables is standard.

**Proposition 3.1.3.** Let $M$ be a prime. Let $u_1, \ldots, u_n$ be independent uniformly distributed random variables in $\mathbb{Z}/MZ$. Let $b_1, \ldots, b_m$ be $(\mathbb{Z}/MZ)$-linear combinations of $u_1, \ldots, u_n$. Then the vector $b = (b_1, \ldots, b_m)$ is uniformly randomly distributed on the range of $b$ in $(\mathbb{Z}/MZ)^m$.

**Proof.** Suppose $b = \sum_j c_j u_j$. So $b = Cu$ where $u = (u_1, \ldots, u_n)$ and $C$ is the matrix with entries $C_{ij} = c_{ij}$. For any $y$ in the image of $C$, the cardinality of the preimage $C^{-1}(y)$ is exactly the cardinality of the kernel of $C$. Indeed, if $Cx = y$, then $C^{-1}(y) = x + \ker(C)$. Since $u$ is uniform, we conclude that $b$ is uniform on the image of $C$. \hfill $\square$

The method of types classifies sequences of symbols according to the relative proportion of occurrences of each symbol. We have used this method before in the proof of the generalized asymptotic sum inequality. In the proof of Theorem 1.2.4 it will play a more important role.

**Definition 3.1.4.** Let $N \in \mathbb{N}$ and let $X$ be a finite “alphabet” set. The type $P_x$ of a sequence $x = (x_1, x_2, \ldots, x_N) \in X^N$ is the relative proportion of occurrences of each symbol of $X$. Type is sometimes called empirical distribution. The possible types for sequences $x \in X^N$ are called the $N$-types on $X$. If $P$ is an $N$-type on $X$, then the set of sequences $x \in X^N$ of type $P$ is called the type class of $P$ and is denoted by $T^N_P$.

**Example 3.1.5.** For $X = \{0, 1\}$ the type of the sequence $x = (1, 0, 0)$ is given by $P_x(0) = 2/3$ and $P_x(1) = 1/3$. The possible 3-types are given by $P(0) = 0$, $P(0) = 1/3$, $P(0) = 2/3$, $P(0) = 1$ (with $P(1) = 1 - P(0)$). The type class of the 3-type given by $P(0) = 2/3$ is $T^3_P = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$.

**Proposition 3.1.6.** Let $N \in \mathbb{N}$ and let $X$ be a finite set. The number of $N$-type classes on $X$ is the binomial coefficient $inom{N + |X| - 1}{|X| - 1} = 2^{\nu(X)}$. Let $P$ be an $N$-type on $X$. The number of sequences in the type class of $P = (p_1, p_2, \ldots)$ is the multinomial coefficient

$$\binom{N}{Np_1, Np_2, \ldots},$$

which is lower bounded by $2^{NH(P) - o(N)}$ and upper bounded by $2^{NH(P)}$. \hfill 27
Proof. The first statement can be proved with the famous stars and bars argument: count the number of ways to arrange $N$ stars and $|X| - 1$ bars in a row. For a proof of the second statement see [CT12, Theorem 12.1.3]. □

3.2. Restriction procedure

We describe a procedure for finding a monomial restriction (see Theorem 1.1.5 for the definition of monomial restriction) from any tensor $\psi$ to $T_r$ for some $r$. Despite its simplicity, this algorithm lies at the heart of the proof of Theorem 1.2.4. Typically, we will apply the algorithm to a tensor $\psi$ of the form $\phi^{\otimes N}$ since we care about the monomial subexponent.

In a graph, a connected component is a maximal connected subgraph. Each vertex of a graph belongs to exactly one connected component.

Lemma 3.2.1. Let $G$ be a graph with $n$ vertices and $m$ edges. Then $G$ has at least $n - m$ connected components.

Proof. A graph without edges has $n$ connected components. For every edge that we add to the graph, we lose at most one connected component. □

Proposition 3.2.2. Let $\psi$ be a tensor in $V_1 \otimes \cdots \otimes V_k$ and let $B_1, \ldots, B_k$ be bases for $V_1, \ldots, V_k$. Let $\Psi$ be the corresponding support of $\psi$. Let $C$ be the set $\{\{a, a'\} \subseteq \Psi \mid a \neq a'; \exists i \, a_i = a'_i\}$. Then $\psi \geq M T_{|\Psi| - |C|}$.

Proof. Let $G$ be the graph with vertex set $\Psi$ and edge set $C$. Let $E(\Psi) \subseteq \Psi$ be a subset of the vertex set that contains exactly one vertex per connected component of $G$. Then $|E(\Psi)| \geq |\Psi| - |C|$ by Theorem 3.2.1. Moreover, because the vertices in $E(\Psi)$ are non-adjacent, there is a monomial restriction $\psi \geq_M T_{|\Psi| - |C|}$. □

We illustrate the procedure in Theorem 3.2.2 with a concrete example where $\psi$ is of the form $\phi^{\otimes 2}$.

Example 3.2.3. Let $D_k = D(1^k)$ be the $k$-tensor $\sum_{\sigma \in S_k} \sigma \cdot b_1 \otimes b_2 \otimes \cdots \otimes b_k$ living in $(\mathbb{C}^k)^{\otimes k}$, where the symmetric group $S_k$ acts by permuting tensor legs. In the standard basis, and identifying $b_i$ with $i$, the tensor $D_k$ has support

$$\Phi = \{(\sigma(1), \sigma(2), \ldots, \sigma(k)) \mid \sigma \in S_k\} \subseteq [k]^k.$$ 

Consider the $N$th power $D_k^{\otimes N} \in (\mathbb{C}^{k^N})^{\otimes k}$. Let $\Phi^{\otimes N} \subseteq ([k]^N)^k$ be its support in the tensor product basis. We write every element in $\Phi^{\otimes N}$ as a $k$-tuple
(I_1, \ldots, I_k) of column N-vectors I_i, so that for k = 3 and N = 2 we get

\[ \Phi^{\otimes 2} = \{\{(0), \{1\}, \{2\}\}, \{(0), \{1\}, \{3\}\}, \{(0), \{2\}, \{3\}\}, \{(1), \{0\}, \{3\}\}, \{(1), \{2\}, \{3\}\}, \{(2), \{0\}, \{3\}\}, \{(2), \{1\}, \{3\}\}\} \]

Order \( \Phi^{\otimes 2} \) as it is written above, and do the procedure as described in Theorem 3.2.2. This yields for example a tensor with support \( \Phi \)

This tensor is isomorphic to \( T_2 \). We thus have \( D_{(1,1,1)}^{\otimes 2} \geq_M T_2 \). Therefore, \( q_M(D_{(1,1,1)}) \geq_M \log_2(2)/2 = 0.5 \). In fact, it is not hard to see that there is a restriction \( D_{(1,1,1)} \geq_M T_2 \), so \( q_M(D_{(1,1,1)}) \geq \log_2(2) = 1 \). The final answer is given by Strassen’s Theorem 1.1.31: \( q_M(D_{(1,1,1)}) = q(D_{(1,1,1)}) = H(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) = \log_2(3) \) which is approximately 1.58496.

### 3.3. Proof of Theorem 1.2.4

We are now ready for the proof of Theorem 1.2.4. In the rest of this section we will use the following notation.

**Notation 3.3.1.** Let \( \phi \) be a \( k \)-tensor, \( B \) a basis, \( \Phi \) the corresponding support. Suppose \( \phi \) is tight with respect to \( \alpha \).

- \( \mathcal{P}_\Phi \) consists of all probability distributions \( P \) on \( \Phi \); and \( \mathbf{P} = (P_1, \ldots, P_k) \) are the marginal distributions of \( P \) on the \( k \) components respectively;

- \( \mathcal{R}_\Phi \) consists of all subsets \( R \subset \Phi \times \Phi \) that are not contained in the diagonal set \( \Delta_\Phi := \{(x, x) \mid x \in \Phi\} \), and such that \( R \subset \mathcal{R}_i \) for some \( i \in \{1, \ldots, k\} \), where \( \mathcal{R}_i := \{(x, y) \in \Phi \times \Phi \mid x_i = y_i\} \).

- \( \mathcal{Q}_{\Phi, \mathbf{P}, \mathcal{R}} \) consists of all probability distributions \( Q \) on \( R \) whose marginals on the \( 2k \) components of \( R \) satisfy \( Q_i = Q_{k+i} = P_i \) for \( i \in \{1, \ldots, k\} \).

Recall that for \( R \subset \Phi \times \Phi \), we defined \( r(R) = r_\alpha(R) \) to be the rank of the matrix with rows \( \{\alpha(x) - \alpha(y) \mid (x, y) \in R\} \) over \( \mathbb{Q} \).

29
Theorem 1.2.4 (Repeated). The monomial subexponent of $\phi$ is at least,

$$q_M(\phi) \geq \max_{P \in \mathcal{P}} \left( H(P) - (k-2) \max_{R \in \mathcal{P}} \frac{\max_{Q \in \mathcal{P}, \mu} H(Q) - H(P)}{r_\alpha(R)} \right).$$

Proof of Theorem 1.2.4. Identify $\Phi$ with $\alpha(\Phi) \subseteq \mathbb{Z}^k$. Let $P = (p_1, p_2, \ldots)$ be a probability distribution on $\Phi$ with rational probabilities $p_i$. Fix a small number $\varepsilon > 0$ and let $N$ be an integer such that every $Np_i$ is an integer (so $P$ is an $N$-type). Let $\mathbf{P} = (P_1, \ldots, P_k)$ be the marginal distributions of $P$ on the $k$ components of $\Phi \subseteq \mathbb{Z}^k$. Let $\Phi^\otimes N$ be the support of $\phi^\otimes N$ in the tensor product basis. Each element in $\Phi$ corresponds to $k$ integers, so each element in $\Phi^\otimes N$ corresponds to $k$ vectors $I_1, I_2, \ldots, I_k \in \mathbb{Z}^N$. (See Theorem 3.2.3.)

Let $\psi$ be the restriction of $\phi^\otimes N$ obtained by keeping only those elements $(I_1, \ldots, I_k)$ in the support for which the type of $I_i$ is $P_i$ for each $i \in [k]$. These include the elements of type $P$. By Theorem 3.1.6, the number of elements remaining can be bounded as

$$|\Phi^\otimes N \cap (T_{P_1}^N \times \cdots \times T_{P_k}^N)| \geq |T_P^N| = \binom{N}{Np_1, Np_2, \ldots} \geq 2^{NH(P) - o(N)},$$

where $T_P^N$ denotes the type class of length $N$ strings of $k$-tuples of integers with type $P$, and $T_{P_i}^N$ denotes the type class of length $N$ strings of integers with type $P_i$. After the restrictions, the new support is

$$\Psi := \Phi^\otimes N \cap (T_{P_1}^N \times \cdots \times T_{P_k}^N), |\Psi| \geq 2^{NH(P) - o(N)}.$$

Let $M$ be a prime between $2^\mu N$ and $2^\mu N$ for some $\mu > 0$ chosen later (this $M$ exists for any $\mu$ and $N$ by Bertrand’s Postulate [AZ14]). Let $B \subseteq \mathbb{N}$ be a $(k-1)$-average-free set with

$$\max(B) < \frac{M}{k-1} \quad \text{and} \quad |B| \geq M^{1-\varepsilon} \quad (19)$$

with $\varepsilon$ as chosen above (such a $B$ exists when $N$ is large enough by Theorem 3.1.2). Let $v_1, \ldots, v_N, u_1, \ldots, u_{k-1}$ be independent uniformly random variables in $\mathbb{Z}/M\mathbb{Z}$, and compute a hash as follows:

$$b_1(I_i) := u_i + \sum_{j=1}^N (I_i)_j v_j \quad \text{for} \ 1 \leq i \leq k-1,$$

$$b_k(I_k) := \frac{1}{k-1} \left( u_1 + \cdots + u_{k-1} - \sum_{j=1}^N (I_k)_j v_j \right),$$

with operations understood in $\mathbb{Z}/M\mathbb{Z}$ (in particular $k-1$ is invertible if $N$ and hence $M$ is large enough). By construction of $b_1, \ldots, b_k$ and since $\phi$ is tight,

$$b_1(I_1) + b_2(I_2) + \cdots + b_{k-1}(I_{k-1}) = (k-1)b_k(I_k) \quad (20)$$

30
holds for every element \((I_1, \ldots, I_k)\) in the support \(\Psi\). Using restrictions again, we keep only those elements such that \(b_i(I_i) \in B\) modulo \(M\) for every \(i \in [k]\). Call the remaining tensor \(\psi'\) and its support \(\Psi'\). Since \(\max(B) < \frac{M}{k}\), the equality in (20) then holds in \(\mathbb{Z}\) for the representatives \(\{0, 1, \ldots, M-1\}\) of \(\mathbb{Z}/M\mathbb{Z}\). By the \((k-1)\)-average-free property of \(B\) this implies \(b_1(I_1) = b_2(I_2) = \cdots = b_k(I_k)\). Summarizing, we restricted \(\psi\) to \(\psi'\) with a monomial restriction in such a way that for every \((I_1, \ldots, I_k) \in \Psi' = \text{supp } \psi'\) we have \(b_1(I_1) = b_2(I_2) = \cdots = b_k(I_k)\).

We apply Theorem 3.2.2 to \(\Psi'\). This yields a monomial restriction

\[
\psi' \geq_M T_{X-Y} \tag{21}
\]

where \(X = |\Psi'|, Y = |\mathcal{C}'|\) for \(\mathcal{C}' := \{(I, I') \in \Psi'^2 \mid I \neq I'; \exists i I_i = I_i'\}\). (The elements of \(\mathcal{C}'\) are ordered pairs. Equation (21) also holds if we take unordered pairs, but ordered pairs will be more convenient later.) We will now find a lower bound on \(\mathbb{E}[X - Y]\) (which will also be a lower bound on the largest possible value of \(X - Y\) where the expectation is taken over the independent uniform choice of the \(u_i\)'s and \(v_i\)'s in \(\mathbb{Z}/M\mathbb{Z}\). We claim to have the following bounds on \(\mathbb{E}[X]\) and \(\mathbb{E}[Y]\):

- \(\mathbb{E}[X] = |B||\Psi|M^{-(k-1)}\) \hspace{1cm} (Claim 1)
- \(\mathbb{E}[Y] \leq |B| \max_{R \in \mathcal{R}_\Phi} \max_{Q \in \mathcal{D}_{\Phi,R}} 2^{NH(Q) + o(N)} M^{-(k-1 + r(R))}\) \hspace{1cm} (Claim 2)

where in the second bound \(\mathcal{R}_\Phi\) and \(\mathcal{D}_{\Phi,R}\) are as in the statement of the theorem. Intuitively, we want to choose \(\mu\) small such that \(\mathbb{E}[X]\) has a large exponent, but we want to choose \(\mu\) large such that \(\mathbb{E}[Y]\) has a smaller exponent than the exponent of \(\mathbb{E}[X]\). To make this precise, we put the claimed bounds together with

- \(|B| \geq M^{1-\epsilon}\)
- \(|\Psi| \geq 2^{NH(P) - o(N)}\)
- \(M \geq 2^{\mu N - o(N)}\)

and use linearity of expectation, to obtain

\[
\mathbb{E}[X - Y] = \mathbb{E}[X] - \mathbb{E}[Y] \\
\geq |B| \left( |\Psi|M^{-(k-1)} - \max_{R \in \mathcal{R}_\Phi} \max_{Q \in \mathcal{D}_{\Phi,R}} 2^{NH(Q) + o(N)} M^{-(k-1 + r(R))} \right) \\
\geq M^{1-\epsilon} 2^{NH(P) - o(N)} M^{-(k-1)} \\
\cdot \left( 1 - \max_{R \in \mathcal{R}_\Phi} \max_{Q \in \mathcal{D}_{\Phi,R}} 2^{NH(Q) - H(P) + o(N)} M^{-r(R)} \right) \\
\geq 2^{NH(P) - (k-2+\epsilon)\mu - o(N)} \\
\cdot \left( 1 - \max_{R \in \mathcal{R}_\Phi} \max_{Q \in \mathcal{D}_{\Phi,R}} 2^{NH(Q) - H(P) - r(R)\mu + o(N)} \right).
\]
We want to choose \( \mu \) small such that the first factor has a large exponent, but we want to choose \( \mu \) large such that the second factor is bounded away from 0. We thus set

\[
\mu = \max_{R \in \mathcal{R}} \max_{Q \in \mathcal{Q}} \frac{H(Q) - H(P) + o(1)}{r(R)}.
\]

After substituting \( \mu \), we get that

\[
q_M(\phi) = 1, \quad \omega_M(\phi, T) = 1, \quad N\omega_M(\phi \otimes N, T) \geq 1, \quad N \log_2 E[X - Y] \geq H(P) - (k - 2 + \epsilon) \max_{R \in \mathcal{R}} \frac{H(Q) - H(P) + o(1)}{r(R)} - o(1).
\]

We note again that here we used the probabilistic method: the largest value of \( X - Y \) is at least \( E[X - Y] \). Finally, we let \( N \to \infty \) and \( \epsilon \to 0 \), and take the supremum over \( P \) with rational entries. By continuity, the latter can be replaced by a maximum over arbitrary real-valued probability distributions \( P \) on \( \Phi \), finishing the main argument of the proof. It remains to prove the two claims.

**Claim 1.** We will prove the claim

\[
E[X] = |\Psi||B| M^{-(k-1)}.
\]

Recall that in the main argument we defined \( \Psi = \Phi \otimes N \cap (T_N^k \times \cdots \times T_N^k) \) and \( \Psi' = \{(I_1, \ldots, I_k) \in \Psi \mid b_i(I_i) \in B \text{ for all } i \in [k]\} \). Recall that for every \( (I_1, \ldots, I_k) \in \Psi' \) we have \( b_1(I_1) = b_2(I_2) = \cdots = b_k(I_k) \). The random variable \( X \) is the size of \( \Psi' \) and can be computed as follows:

\[
E[X] = \sum_{I \in \Psi} \Pr\{b_1(I_1) \in B \land \cdots \land b_k(I_k) \in B\} \cdot 1
= \sum_{I \in \Psi} \sum_{b \in B} \Pr\{b_1(I_1) = b_2(I_2) = \cdots = b_k(I_k) = b\}
= \sum_{I \in \Psi} \sum_{b \in B} \Pr\{b_1(I_1) = b_2(I_2) = \cdots = b_{k-1}(I_{k-1}) = b\}
= \sum_{I \in \Psi} \sum_{b \in B} M^{-(k-1)}
= |\Psi||B| M^{-(k-1)},
\]

using that the random variables \( b_1(I_1), b_2(I_2), \ldots, b_{k-1}(I_{k-1}) \) are independent uniform in \( \mathbb{Z}/M\mathbb{Z} \) because of the presence of \( u_i \) in the definition of \( b_i \).

**Claim 2.** We will prove the claim

\[
E[Y] \leq |B| \max_{R \in \mathcal{R}} \max_{Q \in \mathcal{Q}} 2^{NH(Q) + o(N)} M^{-(k-1+r(R))}.
\]
The number $Y$ is the size of $\mathcal{C}' = \{(I, I') \in \Psi^2 \mid I \neq I'; \exists_i I_i = I'_i\}$, whose expectation can be written in terms of $\mathcal{C} := \{(I, I') \in \Psi^2 \mid I \neq I'; \exists_i I_i = I'_i\}$ as

$$
\mathbb{E}[Y] = \sum_{(I, I') \in \mathcal{C}} \Pr[\forall i \, b_i(I_i) \in B \land b_i(I'_i) \in B]
= \sum_{(I, I') \in \mathcal{C}} \sum_{b \in B} \Pr[b_1(I_1) = \cdots = b_k(I_k) = b_1(I'_1) = \cdots = b_k(I'_k) = b].
$$

Fix a pair $(I_1, \ldots, I_k)$, $(I'_1, \ldots, I'_k) \in \mathcal{C}$. The random variables $b_i(I_i)$ and $b_i(I'_i)$ are linear combinations of independent uniform random variables, and therefore $(b_1(I_1), \ldots, b_k(I_k), b_1(I'_1), \ldots, b_k(I'_k))$ is uniform on the image subspace in $(\mathbb{Z}/M\mathbb{Z})^{2k}$ (see Theorem 3.1.3). This subspace contains $(b, b, \ldots, b)$ for any $b \in \mathbb{Z}/M\mathbb{Z}$, since $u_1 = u_2 = \cdots = u_k = b, v_1 = \cdots = v_N = 0$ is a possible assignment. The probability of the event $(b, b, \ldots, b)$ is thus equal to the reciprocal of the cardinality of the image subspace. This cardinality equals $M$ to the power the rank of the $(N + k - 1) \times 2k$ coefficient matrix

$$
\begin{pmatrix}
1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & \frac{1}{k-1} & \cdots & 1 & \frac{1}{k-1} \\
0 & 1 & 0 & \cdots & 0 & 1 & \cdots & 0 & \frac{1}{k-1} & \cdots & 1 & \frac{1}{k-1} \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 1 & \frac{1}{k-1} & \cdots & \frac{1}{k-1} & \frac{1}{k-1} \\
I_1 & I_2 & \cdots & I_{k-1} & -I_1' & I_2' & \cdots & I_{k-1}' & -\frac{I_k'}{k-1} & \cdots & \frac{I_k'}{k-1} & \frac{I_k'}{k-1}
\end{pmatrix}
$$

over $\mathbb{Z}/M\mathbb{Z}$, with $I_1, \ldots, I_k, I'_1, \ldots, I'_k$ thought of as column vectors. With column and row operations this matrix can be transformed into

$$
\begin{pmatrix}
0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\
I_1 - I_1' & I_2 - I_2' & \cdots & I_{k-1} - I_{k-1}' & I_k - I_k' & 0 & \cdots & 0 & 0
\end{pmatrix}
$$

Here we used that $\sum_{i=1}^k I_i'$ is the zero vector. It follows that the rank of the matrix in (23) is $k - 1$ plus the rank of the $N \times k$ matrix

$$
A(I, I') := \begin{pmatrix}
I_1 - I_1' & I_2 - I_2' & \cdots & I_{k-1} - I_{k-1}' & I_k - I_k'
\end{pmatrix}.
$$

Letting $r_M(I, I')$ denote the rank of $A(I, I')$, equation (22) thus becomes

$$
\mathbb{E}[Y] = \sum_{(I, I') \in \mathcal{C}} \sum_{b \in B} M^{-(k-1+r_M(I, I'))} = |B| \sum_{(I, I') \in \mathcal{C}} M^{-(k-1+r_M(I, I'))}.
$$

(24)
Again fix a pair \(((I_1, \ldots, I_k), (I'_1, \ldots, I'_k)) \in C\). Recall that \(C \subseteq \Psi \times \Phi \subseteq \Phi^{\otimes N} \times \Phi^{\otimes N}\). Therefore, each row in \(A(I, I')\) is of the form \(x - y\) for some \((x, y) \in \Phi \times \Phi\). Namely, the \(i\)th row of \(A(I, I')\) equals \(x - y\) for \(x = ((I_1)_i, (I'_1)_i), \ldots, (I_k)_i\) and \(y = ((I'_1)_i, (I'_1)_i)\). We map every \((I, I') \in C\) to the sequence of pairs corresponding to the rows in \(A(I, I')\),

\[
C \to (\Phi \times \Phi)^N
\]

\[
(I, I') \mapsto (((I_1)_i, \ldots, (I_k)_i), ((I'_1)_i, \ldots, (I'_k)_i))_{i \in [N]}.
\]

The image of this map consists of sequences \(S \in (\Phi \times \Phi)^N\) such that

- there is an \(R \in \mathcal{R}_\Phi\)
- there is an \(N\)-type \(Q \in \mathcal{D}_{\Phi, P, R}\) with \(\text{supp}(Q) = R\)
- \(S \in T^N_Q\).

This map is injective. In words, \(S\) is the sequence of pairs \((x, y)\) corresponding to the rows in the matrix \(A(I, I')\), and \(R\) is the (unique) set of pairs occurring in \(S\), and \(Q\) is the (unique) empirical distribution of the elements of \(R\) in \(S\). Note that the rank of \(A(I, I')\) only depends on the set \(R\). For any \(R \subseteq \Phi \times \Phi\) we thus define \(r_M(R)\) to be the rank of the matrix with rows \(\{x - y \mid (x, y) \in R\}\) over the field \(\mathbb{Z}/M\mathbb{Z}\) and we write \((24)\) as

\[
\mathbb{E}[Y] \leq |B| \sum_{R \in \mathcal{R}_\Phi} \sum_{Q \in \mathcal{D}_{\Phi, P, R}: \text{supp}(Q) = R} \sum_{S \in T^N_Q: Q \text{ is } N\text{-type}} M^{-(k-1+r_M(R))}.
\]

The number of \(N\)-types \(Q\) on \(R\) is at most \(\binom{N+|R|-1}{|R|-1}\), while the size of the type class \(T^N_Q\) is at most \(2^{NH(Q)}\) for any \(Q \in \mathcal{D}_{\Phi, P, R}\) (Theorem 3.1.6). Therefore,

\[
\mathbb{E}[Y] \leq |B| \sum_{R \in \mathcal{R}_\Phi} \binom{N+|R|-1}{|R|-1} \max_{Q \in \mathcal{D}_{\Phi, P, R}} 2^{NH(Q)} M^{-(k-1+r_M(R))}. \tag{25}
\]

Recall that we defined \(r(R)\) as the rank of the matrix with rows \(x - y\) for \((x, y) \in R\), over \(\mathbb{Q}\). The value of \(r_M(R)\) may be less than \(r(R)\), but only for finitely many primes \(M\). Since there are only finitely many possibilities for choosing \(R\), there is an \(N_0\) such that for \(N > N_0\) and for all \(R\) we have \(r_M(R) = r(R)\). In the following we assume that \(N\) (and thus \(M\)) is already large enough for this to hold. In \((25)\), we replace \(r_M(R)\) by \(r(R)\) and upper bound the sum over \(R\) by \(2^{\Phi^2}\) (a constant) times the largest summand, to obtain

\[
\mathbb{E}[Y] \leq |B| \sum_{R \in \mathcal{R}_\Phi} 2^{\Phi^2} \binom{N+|R|-1}{|R|-1} \max_{Q \in \mathcal{D}_{\Phi, P, R}} 2^{NH(Q)} M^{-(k-1+r(R))}.
\]

Now use that the binomial coefficient is polynomial in \(N\). This completes the proof. \(\square\)
3.4. Computational aspects

We use Theorem 3.3.1. Before discussing applications of Theorem 1.2.4, let us focus on how to compute the bound

\[
q_M(\phi) \geq \max_{P \in \mathcal{P}} \left( H(P) - (k - 2) \max_{R \in \mathcal{R}_\Phi} \frac{\max_{Q \in \mathcal{Q}_{P,R}} H(Q) - H(P)}{r_\alpha(R)} \right),
\]

of Theorem 1.2.4.

**Choice of \( Q \).** Given \( P \) and \( R \), computing \( \max_{Q \in \mathcal{Q}_{P,R}} \Phi_{H(Q)} \) is a convex optimization problem (and thus easy) since it amounts to maximizing a concave function over a convex set for each \( i \in [k] \).

**Choice of \( R \).** The set \( \mathcal{R}_\Phi \) may be large, but we can greatly reduce its size as explained in the following two lemmas.

**Lemma 3.4.1.** If \( R, R' \in \mathcal{R}_\Phi \) and \( R \subseteq R' \) and \( r(R) = r(R') \), then

\[
\frac{\max_{Q \in \mathcal{Q}_{P,R}} H(Q)}{r(R)} \leq \frac{\max_{Q \in \mathcal{Q}_{P,R'}} H(Q)}{r(R')}.
\]

**Proof.** We leave the proof to the reader.

Recall that a subset \( R \subseteq \Phi \times \Phi \) is an equivalence relation if for all \( x, y, z \in \Phi \) we have \( (x, x) \in R \); \( (x, y) \in R \Rightarrow (y, x) \in R \); and \( (x, y) \in R \wedge (y, z) \in R \Rightarrow (x, z) \in R \). If \( \Phi = \bigsqcup_i \Phi_i \) is a set partition of \( \Phi \), then \( R = \{(x, y) \in \Phi \times \Phi \mid \exists i : x \in \Phi_i \wedge y \in \Phi_i \} \) is an equivalence relation. The set partition into singletons yields the equality equivalence relation \( \Delta_\Phi = \{(x, x) \mid x \in \Phi \} \).

**Lemma 3.4.2.** Let \( R \in \mathcal{R}_\Phi \). Then there is an \( R' \in \mathcal{R}_\Phi \) which is an equivalence relation and such that \( R \subseteq R' \) and \( r(R) = r(R') \). Therefore, Theorem 1.2.4 still holds if \( \mathcal{R}_\Phi \) is replaced by the set \( \mathcal{R}_\Phi \cap \{\text{equivalence relations}\} \).

**Proof.** We can extend \( R \) to an equivalence relation without increasing its rank. Namely, let \( \Delta_\Phi = \{(x, x) \mid x \in \Phi \} \), \( R^T = \{(y, x) \mid (x, y) \in R\} \) and \( R \circ R = \{(x, z) \mid (x, y), (y, z) \in R\} \). Then \( r(R \cup \Delta_\Phi) = r(R) \), \( r(R \cup R^T) = r(R) \) and \( r(R \cup (R \circ R)) = r(R) \). As mentioned above, if \( R \subseteq R' \) with \( r(R) = r(R') \), then only \( R' \) needs to be considered. Therefore, replacing \( \mathcal{R}_\Phi \) by \( \mathcal{R}_\Phi \cap \{\text{equivalence relations}\} \) will not change the optimal value of the maximization over \( \mathcal{R}_\Phi \).

We state one more simple fact.

**Lemma 3.4.3.** \( r(\mathcal{R}_i) \leq k - 2 \).
Proof. The support $\Phi$ is tight with respect to $\alpha$, so for every $x \in \Phi$ we have $\sum_{j=1}^{k} \alpha_j(x_j) = 0$. So any $w \in \text{Span}_Q\{\alpha(x) - \alpha(y) \mid (x, y) \in \mathcal{R}_i\}$ satisfies the equations $w_i = 0$ and $\sum_{j=1}^{k} w_j = 0$ and hence the span has rank at most $k - 2$.

Choice of $\alpha$. Finally, we discuss the choice of the map $\alpha = (\alpha_1, \ldots, \alpha_k)$. If $\alpha$ and $\beta$ are such maps then $\phi$ is also tight for $\gamma = (\gamma_1, \ldots, \gamma_k)$ with $\gamma_i := \alpha_i + C \beta_i$ for any integer $C$, except that $\gamma_i$ may fail to be injective for finitely many $C$. For a given $R \in \mathcal{R}_\Phi$, $r_\alpha(R)$ is the rank of some matrix $X$ and $r_\beta(R)$ is the rank of some other matrix $Y$, while $r_\gamma(R)$ is the rank of $X + CY$. But the latter is at least $\max\{r_\alpha(R), r_\beta(R)\}$ except for at most finitely many values of $C$. Taking into account every $R$ still excludes only finitely many values of $C$, hence there is at least one good $C$. So in theory one can proceed by finding a new $k$-tuple of injective maps with at least one $r(R)$ higher than what one already has and then improve by finding suitable linear combination. There are finitely many relations and the ranks are never larger than $k - 2$, therefore only finitely many improvements are possible.

3.5. Applications of Theorem 1.2.4

We finish this section by giving some example applications of Theorem 1.2.4, one of which is the result on the weight-$(2, 2)$ Dicke tensor that we use in combination with Theorem 1.2.2 to upper bound the exponent of the complete graph tensor. We use Theorem 3.3.1.

First, as a sanity check, we derive from Theorem 1.2.4 the lower-bound part of Strassen’s Theorem 1.1.31 on the monomial subexponent of tight 3-tensors.

Lemma 3.5.1. Fix marginal distributions $P = (P_1, \ldots, P_k)$ on the $k$ components of $\Phi$ and let $P \in \mathcal{P}_\Phi$ be an element that has maximal entropy among all elements of $\Phi$ with marginals $P_1, \ldots, P_k$. Let $R \in \mathcal{R}_\Phi$ such that $R \subseteq \mathcal{R}_i$ and let $Q \in \mathcal{Q}_{\Phi, P, R}$. Then,

$$H(Q) \leq 2H(P) - H(P_i).$$

Proof. Let $\hat{i} = \{1, \ldots, \hat{i}, \ldots, k\}$ and let $k + \hat{i} = \{k + 1, \ldots, k + \hat{i}, \ldots, 2k\}$, where a hat denotes an omitted index. In this proof, if $K \subseteq [2k]$ is a set of indices, then for any $Q \in \mathcal{Q}_{\Phi, P, R}$ we will write $Q_K$ for the marginal distribution of $Q$ on these indices. Let $(\ast)$ denote the assumption that among all distributions in $\mathcal{P}_\Phi$ with marginals $P_1, \ldots, P_k$ the distribution $P$ is the
one with maximal entropy. Let $Q \in \mathcal{Q}_{\Phi, P, R}$. Then,

$$H(Q) = H(Q_1) + H(Q_{i \cup (k, k+1 \ldots | Q_1)}} \quad \text{(entropy chain rule)}$$

$$\leq H(Q_1) + H(Q | Q_1) + H(Q_{k+1} | Q_1) \quad \text{(strong sub-additivity)}$$

$$\leq H(Q_1) + H(Q_{i \cup (k, k+1 \ldots | Q_1)}} - H(Q_i) \quad \text{(entropy chain rule)}$$

$$\leq H(Q_1) + H(Q_{k, k+1 \ldots | Q_1)}} - H(Q_i) \quad \text{(since $R \subseteq R_i$)}$$

$$\leq 2H(P) - H(P_i), \quad \text{(by (1))}$$

which proves the lemma.

**Corollary 3.5.2 (Strassen [Str91]).** Let $\phi$ be a tight 3-tensor. Then

$$q_{\text{M}}(\phi) \geq \max_{P \in \mathfrak{P}_{\Phi}} \min \{H(P_1), H(P_2), H(P_3)\}.$$  

**Proof.** Let $P \in \mathfrak{P}_{\Phi}$. Then by Theorem 1.2.4,

$$q_{\text{M}}(\phi) \geq H(P) - \max_{R \in \mathfrak{R}_{\Phi}} \frac{\max_{Q \in \mathfrak{Q}_{\Phi, P, R}} H(Q) - H(P)}{r_{\alpha}(R)}. \quad (26)$$

Let $P_1, \ldots, P_k$ be the marginals of $P$. We may assume that $P$ has maximal entropy among all elements of $\mathfrak{P}_{\Phi}$ with marginals $P_1, \ldots, P_k$, since $\mathcal{Q}_{\Phi, P, R}$ depends on $P$ only through its marginals. Let $R \in \mathfrak{R}_{\Phi}$. Then $r(R) \geq 1$ (since $R \not\subseteq \Delta_{\Phi}$) and $r(R) \leq k - 2 = 1$ (Theorem 3.4.3). Therefore, $r(R) = 1$. Combine this fact with (26) and Theorem 3.5.1 to obtain

$$q_{\text{M}}(\phi) \geq H(P) - \max_i (2H(P) - H(P_i) - H(P)) = \min_i H(P_i),$$

which proves the corollary.

Second, we derive from Theorem 1.2.4 the lower-bound part of Theorem 1.1.34 on the subexponent of the W-state tensors $W_k = D_{(1, k-1)}$.

**Corollary 3.5.3 (Vrana–Christandl [VC15]).** Let $k \geq 3$. Then

$$q_{\text{M}}(W_k) = q(W_k) = h(k^{-1}),$$

where $h(p)$ denotes the binary entropy function.

**Proof.** For the proof of the upper bound $q(W_k) \leq h(k^{-1})$ we refer to the proof of Theorem 11 in [VC15]. We will here give a proof for the lower bound $q_{\text{M}}(W_k) \geq h(k^{-1})$, fixing a small gap in the proof of [VC15]. Let $\Phi$ be $\text{supp} W_k$, which we identify with the set $\{(1, 0, \ldots, 0), (0, 1, \ldots, 0), \ldots, (0, 0, \ldots, 1)\}$.
Let \( P \) be the uniform probability distribution on \( \Phi \). Then the marginals of \( P \) are \( P_i(1) = 1 - P_i(0) = \frac{1}{k} \). By Theorem 1.2.4,

\[
\q_k(W_k) \geq \log_2 k - (k - 2) \max_{R \in \mathcal{R}_\Phi} \frac{\max_{Q \in \mathcal{D}_{\Phi,P,R}} H(Q) - \log_2 k}{r(R)}.
\] (27)

We may assume that \( \mathcal{D}_\Phi \) consists of equivalence relations \( R \subseteq \mathcal{R}_1 \) by Theorem 3.4.2 and the symmetry of \( \Phi \). An equivalence relation on \( \Phi \) is just a set partition of \( \Phi \) and we say that the type of such a set partition is the integer partition consisting of the sizes of the parts occurring in the set partition. Up to permuting the elements of \( \Phi \), an equivalence relation on \( \Phi \) is characterized by its type. Let \( \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_{\ell(\lambda)}) \vdash k \) be an integer partition with \( \ell(\lambda) \) parts. Let \( \mathcal{R}_\lambda \subseteq \mathcal{D}_\Phi \) be any equivalence relation of type \( \lambda \). Since \( \mathcal{R}_\lambda \subseteq \mathcal{R}_1 \), we have \( \lambda_{\ell(\lambda)} = 1 \). It is not difficult to see that \( r(\mathcal{R}_\lambda) = k - \ell(\lambda) \) (with \( \alpha \) chosen as in Theorem 1.1.30); indeed if \( C_1, \ldots, C_{\ell(\lambda)} \) are the equivalence classes of \( R_\lambda \) in nonincreasing order, then \( r(C_i) = \lambda_i - 1 \) and \( r(\mathcal{R}_\lambda) = \sum_{i=1}^{\ell(\lambda)} r(C_i) \) since \( \phi = W_k \). Therefore,

\[
r(\mathcal{R}_\lambda) = \sum_{i=1}^{\ell(\lambda)} (\lambda_i - 1) = k - \ell(\lambda).
\] (28)

A probability distribution \( Q \) on \( \mathcal{R}_\lambda \) can be identified with a block-diagonal matrix with \( \lambda_i \times \lambda_i \) blocks, and the condition on the marginals means that the row and column sums are \( \frac{1}{k} \). Using permutation symmetry and concavity of the entropy and permutation symmetry of \( \mathcal{D}_{\Phi,P,R} \), one can see that \( \max_{Q \in \mathcal{D}_{\Phi,P,R}} H(Q) \) is attained when \( Q \) is constant in each block. In this case

\[
H(Q) = \sum_i \lambda_i^2 \left( \frac{1}{\lambda_i k} \log_2 (\lambda_i k) \right) = 2 \log_2 k - H\left( \frac{\lambda}{k} \right).
\] (29)

Combine (27), (28) and (29) to get

\[
\q_k(W_k) \geq \log_2 k - (k - 2) \max_{\lambda} \frac{\log_2 k - H\left( \frac{\lambda}{k} \right)}{k - \ell(\lambda)}.
\]

If \( k \) and \( \ell = \ell(\lambda) \) are fixed, then the minimum of \( H\left( \frac{\lambda}{k} \right) \) is attained at the partition \( \lambda = (k - \ell + 1, 1^{\ell-1}) \), because the other distributions can be expressed as a convex combinations of its permutations. Therefore we need to look for the maximum of

\[
\frac{\log_2 k - H\left( \frac{k-\ell+1,1^{\ell-1}}{k} \right)}{k - \ell} = \frac{(k - \ell + 1) \log_2 (k - \ell + 1)}{k(k - \ell)}.
\]

Taking the derivative with respect to \( \ell \), we get

\[
\frac{d}{d\ell} \frac{(k - \ell + 1) \log_2 (k - \ell + 1)}{k(k - \ell)} = \frac{\ln(1 + k - \ell) - (k - \ell)}{k(k - \ell)^2 \ln 2} \leq 0,
\]

38
Then \( q(2) \) is the Dicke tensor, with the two types which we identify with which proves the corollary.

**Proof.** Let \( R \) be any equivalence relations of type \( (2,2) \) and \( S \) are equivalent. There is only one such equivalence relation, namely uniform distributions \((1,1,0,0), (1,0,1,0), \ldots \). Then \( H(P) = \log_2 6 \). Note that the marginals \( P_i \) are uniform distributions on \( \{0,1\} \). Let \( R \subseteq \mathcal{R}_\Phi \). By the permutation symmetry of \( D_{(2,2)} \), we may assume that

\[
R \subseteq \mathcal{R}_1 = \{(1,1,0,0), (1,0,1,0), (1,0,0,1)\}^2 \cup \{(0,0,1,1), (0,1,0,1), (0,1,1,0)\}^2.
\]

By Theorem 3.4.2, we may assume that \( R \) is an equivalence relation. Define the set

\[
S = \{(1,1,0,0), (1,0,1,0), (1,0,0,1)\}.
\]

If \((x,y) \in R\), then \( R \subseteq R' := R \cup \{(1,1,1,1) - x, (1,1,1,1) - y\} \in \mathcal{R}_\Phi \) and \( r(R) = r(R') \), and hence we may assume that if \((x,y) \in R\) then also \((1,1,1,1) - x, (1,1,1,1) - y\) \(\in R\) (Theorem 3.4.1). We thus restrict ourselves to the equivalence relations \( R \subseteq \mathcal{R}_1 \) of the following three types. In type (3) all three elements of \( S \) are equivalent. There is only one such equivalence relation, namely the whole set \( \mathcal{R}_1 \), which has size 18. In type (2,1), two elements of \( S \) are mutually equivalent and inequivalent to the third element (which is equivalent to itself). Then the size of \( R \) is 10. In type (1,1,1), all elements of \( S \) are inequivalent. However, this means that \( R \) is contained in the diagonal \( \Delta_\Phi = \{(x,x) \mid x \in \Phi\} \). Such an \( R \) is not feasible. So we are left with the two types (3) and (2,1). For type (3) we have \( r(R) = 2 \) while for type (2,1) we have \( r(R) = 1 \). In both cases, the uniform probability distribution \( Q \) on \( R \) has marginals \( P_1, P_2, P_3, P_4, P_5, P_2, P_3, P_4 \), namely uniform distributions on \( \{0,1\} \). Therefore, this \( Q \) is the optimal \( Q \) and \( H(Q) = \log_2 |R| \). Letting \( R_{(3)} \) and \( R_{(2,1)} \) be any equivalence relations of type (3) and (2,1) respectively,
we obtain

\[ q_{\text{M}}(D_{(2,2)}) \geq \min \{ H(P) - \frac{2 \log_2 |R_{(3)}|}{r(R_{(3)})} - H(P), \]

\[ H(P) - \frac{2 \log_2 |R_{(2,1)}|}{r(R_{(2,1)})} - H(P) \} \]

\[ = \min \{ \log_2 6 - \frac{2}{3} (\log_2 18 - \log_2 6), \]

\[ \log_2 6 - \frac{2}{3} (\log_2 10 - \log_2 6) \}\]

\[ = \min \{ 1, \log_2 \frac{54}{25} \} = 1 \]

On the other hand, \( D_{(2,2)} \) has a flattening of rank 2, so also the upper bound \( q(D_{(2,2)}) \leq 1 \) holds.

\[ \square \]

**Acknowledgements.** The authors wish to thank Peter Bürgisser for helpful discussions. MC acknowledges financial support from the European Research Council (ERC Grant Agreement no. 337603), the Danish Council for Independent Research (Sapere Aude), and VILLUM FONDEN via the QMATH Centre of Excellence (Grant no. 10059). JZ is supported by NWO through the research programme 617.023.116.

**References**


