Powers of Tensors and Fast Matrix Multiplication

François Le Gall

Department of Computer Science
Graduate School of Information Science and Technology
The University of Tokyo

Simons Institute, 12 November 2014
Overview of our Results
Algebraic Complexity of Matrix Multiplication

Compute the product of two $n \times n$ matrices $A$ and $B$ over a field $\mathbb{F}$.

- **Model:** algebraic circuits
  - **gates:** $+, -, \times, \div$ (operations on two elements of the field)
  - **input:** $a_{ij}, b_{ij}$ ($2n^2$ inputs)
  - **output:** $c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$ ($n^2$ outputs)

$C_M(n) =$ minimal number of algebraic operations needed to compute the product

$\omega = \inf \left\{ \alpha \mid C_M(n) \leq n^\alpha \text{ for all large enough } n \right\}$

Obviously, $2 \leq \omega \leq 3$. 

Note: may depend on the field $\mathbb{F}$.
History of the main improvements on the exponent of square matrix multiplication

<table>
<thead>
<tr>
<th>Upper bound</th>
<th>Year</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega \leq 3$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega &lt; 2.81$</td>
<td>1969</td>
<td>Strassen</td>
</tr>
<tr>
<td>$\omega &lt; 2.79$</td>
<td>1979</td>
<td>Pan</td>
</tr>
<tr>
<td>$\omega &lt; 2.78$</td>
<td>1979</td>
<td>Bini, Capovani, Romani and Lotti</td>
</tr>
<tr>
<td>$\omega &lt; 2.55$</td>
<td>1981</td>
<td>Schönhage</td>
</tr>
<tr>
<td>$\omega &lt; 2.53$</td>
<td>1981</td>
<td>Pan</td>
</tr>
<tr>
<td>$\omega &lt; 2.52$</td>
<td>1982</td>
<td>Romani</td>
</tr>
<tr>
<td>$\omega &lt; 2.50$</td>
<td>1982</td>
<td>Coppersmith and Winograd</td>
</tr>
<tr>
<td>$\omega &lt; 2.48$</td>
<td>1986</td>
<td>Strassen</td>
</tr>
<tr>
<td>$\omega &lt; 2.376$</td>
<td>1987</td>
<td>Coppersmith and Winograd</td>
</tr>
<tr>
<td>$\omega &lt; 2.373$</td>
<td>2010</td>
<td>Stothers</td>
</tr>
<tr>
<td>$\omega &lt; 2.3729$</td>
<td>2012</td>
<td>Vassilevska Williams</td>
</tr>
<tr>
<td>$\omega &lt; 2.3728639$</td>
<td>2014</td>
<td>Le Gall</td>
</tr>
</tbody>
</table>
## History of the main improvements on the exponent of square matrix multiplication

<table>
<thead>
<tr>
<th>Upper bound</th>
<th>Year</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega \leq 3$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega &lt; 2.81$</td>
<td>1969</td>
<td>Strassen</td>
</tr>
<tr>
<td>$\omega &lt; 2.79$</td>
<td>1979</td>
<td>Pan</td>
</tr>
<tr>
<td>$\omega &lt; 2.78$</td>
<td>1979</td>
<td>Bini, Capovani, Romani and Lotti</td>
</tr>
<tr>
<td>$\omega &lt; 2.55$</td>
<td>1981</td>
<td>Schöhnhage</td>
</tr>
<tr>
<td>$\omega &lt; 2.53$</td>
<td>1981</td>
<td>Pan</td>
</tr>
<tr>
<td>$\omega &lt; 2.52$</td>
<td>1982</td>
<td>Romani</td>
</tr>
<tr>
<td>$\omega &lt; 2.50$</td>
<td>1982</td>
<td>Coppersmith and Winograd</td>
</tr>
<tr>
<td>$\omega &lt; 2.48$</td>
<td>1986</td>
<td>Strassen</td>
</tr>
<tr>
<td>$\omega &lt; 2.376$</td>
<td>1987</td>
<td>Coppersmith and Winograd</td>
</tr>
<tr>
<td>$\omega &lt; 2.373$</td>
<td>2010</td>
<td>Stothers</td>
</tr>
<tr>
<td>$\omega &lt; 2.3729$</td>
<td>2012</td>
<td>Vassilevska Williams</td>
</tr>
<tr>
<td>$\omega &lt; 2.3728639$</td>
<td>2014</td>
<td>Le Gall</td>
</tr>
</tbody>
</table>
History of the main improvements on the exponent of square matrix multiplication

| ω < 2.48 | 1986 | Strassen | LM-based analysis v1 |
| ω < 2.376 | 1987 | Coppersmith and Winograd | LM-based analysis v2.0 |
| ω < 2.373 | 2010 | Stothers | LM-based analysis v2.1 |
| ω < 2.3729 | 2012 | Vassilevska Williams | LM-based analysis v2.2 |
| ω < 2.3728639 | 2014 | Le Gall | LM-based analysis v2.3 |

The tensors considered become more difficult to analyze (technical difficulties appear + the “size” of the tensor increases)

Previous versions (up to v2.2):

analyzing the tensor required solving a complicated optimization problem (difficult when the size of the tensor increases)

Our new technique (v2.3):

analyzing the tensor (i.e., obtaining an upper bound on ω from it) can be done in time polynomial in the size of the tensor

- analysis based on convex optimization
Applications of our method

any tensor from which an upper bound on $\omega$ can be obtained from the laser method

Laser-method-based analysis v2.3

corresponding upper bound on $\omega$

which tensor? powers of the basic tensor from Coppersmith and Winograd’s paper

analysis of the $m$-th power of the tensor by CW

<table>
<thead>
<tr>
<th>$m$</th>
<th>Upper bound</th>
<th>Number of variables in the optimization problem</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\omega &lt; 2.3871900$</td>
<td>1</td>
<td>CW (1987)</td>
</tr>
<tr>
<td>2</td>
<td>$\omega &lt; 2.3754770$</td>
<td>3</td>
<td>CW (1987)</td>
</tr>
</tbody>
</table>

$\omega < 2.48$ | 1986 | Strassen | LM-based analysis v1

$\omega < 2.376$ | 1987 | Coppersmith and Winograd | LM-based analysis v2.0
Applications of our method

any tensor from which an upper bound on $\omega$ can be obtained from the laser method

Laser-method-based analysis v2.3 polynomial time

corresponding upper bound on $\omega$

which tensor? powers of the basic tensor from Coppersmith and Winograd’s paper

analysis of the $m$-th power of the tensor by CW

<table>
<thead>
<tr>
<th>$m$</th>
<th>Upper bound</th>
<th>Number of variables in the optimization problem</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\omega &lt; 2.3871900$</td>
<td>1</td>
<td>CW (1987)</td>
</tr>
<tr>
<td>2</td>
<td>$\omega &lt; 2.3754770$</td>
<td>3</td>
<td>CW (1987)</td>
</tr>
<tr>
<td>4</td>
<td>$\omega &lt; 2.3729269$</td>
<td>9</td>
<td>Stothers (2010)</td>
</tr>
</tbody>
</table>

$\omega < 2.48$ 1986 Strassen LM-based analysis v1
$\omega < 2.376$ 1987 Coppersmith and Winograd LM-based analysis v2.0
$\omega < 2.373$ 2010 Stothers LM-based analysis v2.1
Applications of our method

any tensor from which an upper bound on $\omega$ can be obtained from the laser method

Laser-method-based analysis v2.3
polynomial time

corresponding upper bound on $\omega$

which tensor? powers of the basic tensor from Coppersmith and Winograd’s paper

analysis of the $m$-th power of the tensor by CW

<table>
<thead>
<tr>
<th>$m$</th>
<th>Upper bound</th>
<th>Number of variables in the optimization problem</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\omega &lt; 2.3871900$</td>
<td>1</td>
<td>CW (1987)</td>
</tr>
<tr>
<td>2</td>
<td>$\omega &lt; 2.3754770$</td>
<td>3</td>
<td>CW (1987)</td>
</tr>
<tr>
<td>4</td>
<td>$\omega &lt; 2.3729269$</td>
<td>9</td>
<td>Stothers (2010)</td>
</tr>
<tr>
<td>8</td>
<td>$\omega &lt; 2.3729$</td>
<td>29</td>
<td>Vassilevska Williams (2012)</td>
</tr>
</tbody>
</table>

$\omega < 2.48$ 1986 Strassen LM-based analysis v1
$\omega < 2.376$ 1987 Coppersmith and Winograd LM-based analysis v2.0
$\omega < 2.373$ 2010 Stothers LM-based analysis v2.1
$\omega < 2.3729$ 2012 Vassilevska Williams LM-based analysis v2.2
**Applications of our method**

Any tensor from which an upper bound on $\omega$ can be obtained from the laser method.

Laser-method-based analysis v2.3 polynomial time.

Corresponding upper bound on $\omega$.

Which tensor? Powers of the basic tensor from Coppersmith and Winograd’s paper.

Analysis of the $m$-th power of the tensor by CW.

<table>
<thead>
<tr>
<th>$m$</th>
<th>Upper bound</th>
<th>Number of variables in the optimization problem</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\omega &lt; 2.3871900$</td>
<td>1</td>
<td>CW (1987)</td>
</tr>
<tr>
<td>2</td>
<td>$\omega &lt; 2.3754770$</td>
<td>3</td>
<td>CW (1987)</td>
</tr>
<tr>
<td>4</td>
<td>$\omega &lt; 2.3729269$</td>
<td>9</td>
<td>Stothers (2010)</td>
</tr>
<tr>
<td>8</td>
<td>$\omega &lt; 2.3729$</td>
<td>29</td>
<td>Vassilevska Williams (2012)</td>
</tr>
<tr>
<td>16</td>
<td>$\omega &lt; 2.3728640$</td>
<td>101</td>
<td>Le Gall (2014)</td>
</tr>
<tr>
<td>32</td>
<td>$\omega &lt; 2.3728639$</td>
<td>373</td>
<td>Le Gall (2014)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\omega$</th>
<th>Year</th>
<th>Authors</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>$&lt; 2.48$</td>
<td>1986</td>
<td>Strassen</td>
<td>LM-based analysis v1</td>
</tr>
<tr>
<td>$&lt; 2.376$</td>
<td>1987</td>
<td>Coppersmith and Winograd</td>
<td>LM-based analysis v2.0</td>
</tr>
<tr>
<td>$&lt; 2.373$</td>
<td>2010</td>
<td>Stothers</td>
<td>LM-based analysis v2.1</td>
</tr>
<tr>
<td>$&lt; 2.3729$</td>
<td>2012</td>
<td>Vassilevska Williams</td>
<td>LM-based analysis v2.2</td>
</tr>
<tr>
<td>$&lt; 2.3728639$</td>
<td>2014</td>
<td>Le Gall</td>
<td>LM-based analysis v2.3</td>
</tr>
</tbody>
</table>
How to Obtain Upper Bounds on $\omega$?
Strassen’s algorithm (for the product of two 2x2 matrices)

Goal: compute the product of \( A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \) by \( B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \)

1. Compute:
   
   \[
   \begin{align*}
   m_1 &= a_{11} \times (b_{12} - b_{22}), \\
   m_2 &= (a_{11} + a_{12}) \times b_{22}, \\
   m_3 &= (a_{21} + a_{22}) \times b_{11}, \\
   m_4 &= a_{22} \times (b_{21} - b_{11}), \\
   m_5 &= (a_{11} + a_{22}) \times (b_{11} + b_{22}), \\
   m_6 &= (a_{12} - a_{22}) \times (b_{21} + b_{22}), \\
   m_7 &= (a_{11} - a_{21}) \times (b_{11} + b_{12})..
   \end{align*}
   
2. Output:
   
   \[
   \begin{align*}
   -m_2 + m_4 + m_5 + m_6 &= c_{11}, \\
   m_1 + m_2 &= c_{12}, \\
   m_3 + m_4 &= c_{21}, \\
   m_1 - m_3 + m_5 - m_7 &= c_{22}.
   \end{align*}
   
7 multiplications  18 additions/subtractions
Strassen’s algorithm (for the product of two $2^k \times 2^k$ matrices)

Goal: compute the product of

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \text{ by } B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

1. Compute:

$$m_1 = a_{11} \ast (b_{12} - b_{22}),$$
$$m_2 = (a_{11} + a_{12}) \ast b_{22},$$
$$m_3 = (a_{21} + a_{22}) \ast b_{11},$$
$$m_4 = a_{22} \ast (b_{21} - b_{11}),$$
$$m_5 = (a_{11} + a_{22}) \ast (b_{11} + b_{22}),$$
$$m_6 = (a_{12} - a_{22}) \ast (b_{21} + b_{22}),$$
$$m_7 = (a_{11} - a_{21}) \ast (b_{11} + b_{12}).$$

2. Output:

$$-m_2 + m_4 + m_5 + m_6 = c_{11},$$
$$m_1 + m_2 = c_{12},$$
$$m_3 + m_4 = c_{21},$$
$$m_1 - m_3 + m_5 - m_7 = c_{22}.$$

7 multiplications 18 additions/subtractions

Recursive application gives

$$C_M(2^k) = O(7^k) = O((2^k)^{\log_2 7})$$

$$\implies \omega \leq \log_2(7) = 2.807\ldots \quad \text{[Strassen 69]}$$
Strassen’s algorithm (for the product of two $2^k \times 2^k$ matrices)

More generally:

Suppose that the product of two $m \times m$ matrices can be computed with $t$ multiplications. Then

$$\omega \leq \log_m(t) \text{ or, equivalently, } m^{\omega} \leq t.$$  

Strassen’s algorithm is the case $m = 2$ and $t = 7$

7 multiplications 18 additions/subtractions

Recursive application gives

$$C_M(2^k) = O(7^k) = O((2^k)^{\log_2 7})$$

$$\implies \omega \leq \log_2(7) = 2.807...$$  

[Strassen 69]
The tensor of matrix multiplication

**Definition**

The tensor corresponding to the multiplication of an $m \times n$ matrix by an $n \times p$ matrix is

$$\langle m, n, p \rangle = \sum_{i=1}^{m} \sum_{j=1}^{p} \sum_{k=1}^{n} a_{ik} \otimes b_{kj} \otimes c_{ij}.$$  

**Intuitive interpretation:**

- this is a formal sum
- when the $a_{ik}$ and the $b_{kj}$ are replaced by the corresponding entries of matrices, the coefficient of $c_{ij}$ becomes $\sum_{k=1}^{n} a_{ik} b_{kj}$
General 3-tensors

Consider three vector spaces $U$, $V$ and $W$ over $\mathbb{F}$

Take bases of $U$, $V$ and $W$:

\[ U = \text{span}\{x_1, \ldots, x_{\text{dim}(U)}\} \]
\[ V = \text{span}\{y_1, \ldots, y_{\text{dim}(V)}\} \]
\[ W = \text{span}\{z_1, \ldots, z_{\text{dim}(W)}\} \]

A tensor over $(U, V, W)$ is an element of $U \otimes V \otimes W$

i.e., a formal sum

\[
T = \sum_{u=1}^{\text{dim}(U)} \sum_{v=1}^{\text{dim}(V)} \sum_{w=1}^{\text{dim}(W)} d_{uvw} x_u \otimes y_v \otimes z_w \in \mathbb{F}
\]

"a three-dimension array with $\text{dim}(U) \times \text{dim}(V) \times \text{dim}(W)$ entries in $\mathbb{F}$"
General 3-tensors

A tensor over \((U, V, W)\) is an element of \(U \otimes V \otimes W\), i.e., a formal sum:

\[
T = \sum_{u=1}^{\dim(U)} \sum_{v=1}^{\dim(V)} \sum_{w=1}^{\dim(W)} d_{uvw} x_u \otimes y_v \otimes z_w
\]

where

- \(\dim(U) = mn\), \(\dim(V) = np\) and \(\dim(W) = mp\)
- \(U = \text{span}\{\{a_{ik}\}_{1 \leq i \leq m, 1 \leq k \leq n}\}\)
- \(V = \text{span}\{\{b_{k'j}\}_{1 \leq k' \leq n, 1 \leq j \leq p}\}\)
- \(W = \text{span}\{\{c_{i'j'}\}_{1 \leq i' \leq m, 1 \leq j' \leq p}\}\)

\[
d_{ikk'j'j} = \begin{cases} 
1 & \text{if } i = i', j = j', k = k' \\
0 & \text{otherwise}
\end{cases}
\]

Definition

The tensor corresponding to the multiplication of an \(m \times n\) matrix by an \(n \times p\) matrix is:

\[
\langle m, n, p \rangle = \sum_{i=1}^{m} \sum_{j=1}^{p} \sum_{k=1}^{n} a_{ik} \otimes b_{kj} \otimes c_{ij}.
\]
**Rank**

**Definition**

The tensor corresponding to the multiplication of an $m \times n$ matrix by an $n \times p$ matrix is

$$\langle m, n, p \rangle = \sum_{i=1}^{m} \sum_{j=1}^{p} \sum_{k=1}^{n} a_{ik} \otimes b_{kj} \otimes c_{ij}.$$ 

$$R(\langle m, n, p \rangle) \leq mnp$$

$$\langle 2, 2, 2 \rangle = a_{11} \otimes (b_{12} - b_{22}) \otimes (c_{12} + c_{22}) + (a_{11} + a_{12}) \otimes b_{22} \otimes (-c_{11} + c_{12}) + (a_{21} + a_{22}) \otimes b_{11} \otimes (c_{21} - c_{22}) + a_{22} \otimes (b_{21} - b_{11}) \otimes (c_{11} + c_{21}) + (a_{11} + a_{22}) \otimes (b_{11} + b_{22}) \otimes (c_{11} + c_{22}) + (a_{12} - a_{22}) \otimes (b_{21} + b_{22}) \otimes c_{11} + (a_{11} - a_{21}) \otimes (b_{11} + b_{12}) \otimes (-c_{22})$$

Strassen’s algorithm gives

$$R(\langle 2, 2, 2 \rangle) \leq 7$$

rank = # of multiplications of the best (bilinear) algorithm
How to obtain upper bounds on $\omega$?

Remember:

Suppose that the product of two $m \times m$ matrices can be computed with $t$ multiplications. Then

$$\omega \leq \log_m(t) \text{ or, equivalently, } m^\omega \leq t.$$ 

In our terminology: $R(\langle m, m, m \rangle) \leq t \implies m^\omega \leq t$

<table>
<thead>
<tr>
<th>First generalization:</th>
<th>Theorem</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R(\langle m, n, p \rangle) \leq t \implies (mnp)^{\omega/3} \leq t$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Second generalization:</th>
<th>Theorem</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Bini et al. 1979]</td>
<td>$R(\langle m, n, p \rangle) \leq t \implies (mnp)^{\omega/3} \leq t$</td>
</tr>
</tbody>
</table>

border rank $R(\langle m, n, p \rangle) \leq R(\langle m, n, p \rangle)$
How to obtain upper bounds on $\omega$?

Third generalization:

Theorem (the asymptotic sum inequality, special case) [Schönhage 1981]

$$R(\langle m_1, n_1, p_1 \rangle \oplus \langle m_2, n_2, p_2 \rangle) \leq t \implies (m_1 n_1 p_1)^{\omega/3} + (m_2 n_2 p_2)^{\omega/3} \leq t$$

Direct sum

Theorem (the asymptotic sum inequality, general form) [Schönhage 1981]

$$R\left(\bigoplus_{i=1}^{k} \langle m_i, n_i, p_i \rangle\right) \leq t \implies \sum_{i=1}^{k} (m_i n_i p_i)^{\omega/3} \leq t$$

First generalization:

Theorem

$$R(\langle m, n, p \rangle) \leq t \implies (mnp)^{\omega/3} \leq t$$

Second generalization: [Bini et al. 1979]

Theorem

$$R(\langle m, n, p \rangle) \leq t \implies (mnp)^{\omega/3} \leq t$$

Border rank

$$R(\langle m, n, p \rangle) \leq R(\langle m, n, p \rangle)$$
History of the main improvements on the exponent of square matrix multiplication

<table>
<thead>
<tr>
<th>Upper bound</th>
<th>Year</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega \leq 3$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega &lt; 2.81$</td>
<td>1969</td>
<td>Strassen</td>
</tr>
<tr>
<td>$\omega &lt; 2.79$</td>
<td>1979</td>
<td>Pan</td>
</tr>
<tr>
<td>$\omega &lt; 2.78$</td>
<td>1979</td>
<td>Bini et al.</td>
</tr>
<tr>
<td>$\omega &lt; 2.55$</td>
<td>1981</td>
<td>Schönhage</td>
</tr>
<tr>
<td>$\omega &lt; 2.53$</td>
<td>1981</td>
<td>Pan</td>
</tr>
<tr>
<td>$\omega &lt; 2.52$</td>
<td>1982</td>
<td>Romani</td>
</tr>
<tr>
<td>$\omega &lt; 2.50$</td>
<td>1982</td>
<td>Coppersmith and Winograd</td>
</tr>
<tr>
<td>$\omega &lt; 2.48$</td>
<td>1986</td>
<td>Strassen</td>
</tr>
<tr>
<td>$\omega &lt; 2.376$</td>
<td>1987</td>
<td>Coppersmith and Winograd</td>
</tr>
<tr>
<td>$\omega &lt; 2.373$</td>
<td>2010</td>
<td>Stothers</td>
</tr>
<tr>
<td>$\omega &lt; 2.3729$</td>
<td>2012</td>
<td>Vassilevska Williams</td>
</tr>
<tr>
<td>$\omega &lt; 2.3728639$</td>
<td>2014</td>
<td>Le Gall</td>
</tr>
</tbody>
</table>

- Upper bound on $\omega$ from the analysis of the rank of a tensor
- Analysis of the border rank of a tensor
- Analysis of a tensor by the asymptotic sum inequality
- Analysis of a tensor by the laser method
The Laser Method on a Simpler Example
Why this is called the “laser method”?

limited by our ignorance about $\omega$. Surprisingly, the exact knowledge of the left end of $\Delta_c$ can be used to obtain an improved estimate for its right end, namely $\omega < 2.48$. The method employed is called laser method [27], since it is reminiscent of the generation of coherent light.

from V. Strassen.  
*Algebra and Complexity.* 

<table>
<thead>
<tr>
<th>Upper bound</th>
<th>Year</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega \leq 3$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega &lt; 2.81$</td>
<td>1969</td>
<td>Strassen</td>
</tr>
<tr>
<td>$\omega &lt; 2.79$</td>
<td>1979</td>
<td>Pan</td>
</tr>
<tr>
<td>$\omega &lt; 2.78$</td>
<td>1979</td>
<td>Bini, Capovani, Romani and Lotti</td>
</tr>
<tr>
<td>$\omega &lt; 2.55$</td>
<td>1981</td>
<td>Schönhage</td>
</tr>
<tr>
<td>$\omega &lt; 2.53$</td>
<td>1981</td>
<td>Pan</td>
</tr>
<tr>
<td>$\omega &lt; 2.52$</td>
<td>1982</td>
<td>Romani</td>
</tr>
<tr>
<td>$\omega &lt; 2.50$</td>
<td>1982</td>
<td>Coppersmith and Winograd</td>
</tr>
<tr>
<td>$\omega &lt; 2.48$</td>
<td>1986</td>
<td>Strassen</td>
</tr>
<tr>
<td>$\omega &lt; 2.376$</td>
<td>1987</td>
<td>Coppersmith and Winograd</td>
</tr>
<tr>
<td>$\omega &lt; 2.373$</td>
<td>2010</td>
<td>Stothers</td>
</tr>
<tr>
<td>$\omega &lt; 2.3729$</td>
<td>2012</td>
<td>Vassilevska Williams</td>
</tr>
<tr>
<td>$\omega &lt; 2.3728639$</td>
<td>2014</td>
<td>Le Gall</td>
</tr>
</tbody>
</table>

Ref. [27] variants (improvements) of the laser method
The first CW construction

Let $q$ be a positive integer.

Consider three vector spaces $U$, $V$ and $W$ of dimension $q + 1$ over $\mathbb{F}$.

$$U = \text{span}\{x_0, \ldots, x_q\}$$
$$V = \text{span}\{y_0, \ldots, y_q\} \quad W = \text{span}\{z_0, \ldots, z_q\}$$

Coppersmith and Winograd (1987) introduced the following tensor:

$$T_{\text{easy}} = T_{\text{easy}}^{011} + T_{\text{easy}}^{101} + T_{\text{easy}}^{110},$$

tensor over $(U, V, W)$

where

$$T_{\text{easy}}^{011} = \sum_{i=1}^{q} x_0 \otimes y_i \otimes z_i \cong \langle 1, 1, q \rangle$$
$$T_{\text{easy}}^{101} = \sum_{i=1}^{q} x_i \otimes y_0 \otimes z_i \cong \langle q, 1, 1 \rangle$$
$$T_{\text{easy}}^{110} = \sum_{i=1}^{q} x_i \otimes y_i \otimes z_0 \cong \langle 1, q, 1 \rangle$$

$$T_{\text{easy}}^{011} = \sum_{i=1}^{q} x_{0i} \otimes y_{0i} \otimes z_{0i} \quad 1 \times 1 \text{ matrix by } 1 \times q \text{ matrix}$$
$$T_{\text{easy}}^{101} = \sum_{i=1}^{q} x_{i0} \otimes y_{00} \otimes z_{i0} \quad q \times 1 \text{ matrix by } 1 \times 1 \text{ matrix}$$
$$T_{\text{easy}}^{110} = \sum_{i=1}^{q} x_{0i} \otimes y_{i0} \otimes z_{00} \quad 1 \times q \text{ matrix by } q \times 1 \text{ matrix}$$
The first CW construction

\[ U = \text{span}\{x_0, \ldots, x_q\} \]
\[ V = \text{span}\{y_0, \ldots, y_q\} \quad W = \text{span}\{z_0, \ldots, z_q\} \]

\[ U = U_0 \oplus U_1, \quad \text{where} \quad U_0 = \text{span}\{x_0\} \quad \text{and} \quad U_1 = \text{span}\{x_1, \ldots, x_q\} \]

\[ V = V_0 \oplus V_1, \quad \text{where} \quad V_0 = \text{span}\{y_0\} \quad \text{and} \quad V_1 = \text{span}\{y_1, \ldots, y_q\} \]

\[ W = W_0 \oplus W_1, \quad \text{where} \quad W_0 = \text{span}\{z_0\} \quad \text{and} \quad W_1 = \text{span}\{z_1, \ldots, z_q\} \]

Coppersmith and Winograd (1987) introduced the following tensor:

\[ T_{\text{easy}} = T_{\text{easy}}^{011} + T_{\text{easy}}^{101} + T_{\text{easy}}^{110}, \]

This is not a direct sum

where

\[ T_{\text{easy}}^{011} = \sum_{i=1}^{q} x_0 \otimes y_i \otimes z_i \]

\[ T_{\text{easy}}^{101} = \sum_{i=1}^{q} x_i \otimes y_0 \otimes z_i \]

\[ T_{\text{easy}}^{110} = \sum_{i=1}^{q} x_i \otimes y_i \otimes z_0 \]

tensor over \((U_0, V_1, W_1)\)  
tensor over \((U_1, V_0, W_1)\)  
tensor over \((U_1, V_1, W_0)\)
The first CW construction

\[ T_{\text{easy}} = T_{\text{easy}}^{011} + T_{\text{easy}}^{101} + T_{\text{easy}}^{110} \]

\[ R(T_{\text{easy}}) \leq q + 2 \]

Actually, \( R(T_{\text{easy}}) = q + 2 \)

Since the sum is not direct, we cannot use the asymptotic sum inequality.

Consider \( T_{\text{easy}}^{\otimes 2} = (T_{\text{easy}}^{011} + T_{\text{easy}}^{101} + T_{\text{easy}}^{110}) \otimes (T_{\text{easy}}^{011} + T_{\text{easy}}^{101} + T_{\text{easy}}^{110}) \)

\[ = T_{\text{easy}}^{011} \otimes T_{\text{easy}}^{011} + T_{\text{easy}}^{011} \otimes T_{\text{easy}}^{101} + \cdots + T_{\text{easy}}^{110} \otimes T_{\text{easy}}^{110} \quad \text{(9 terms)} \]

Consider \( T_{\text{easy}}^{\otimes N} = T_{\text{easy}}^{011} \otimes \cdots \otimes T_{\text{easy}}^{011} + \cdots + T_{\text{easy}}^{110} \otimes \cdots \otimes T_{\text{easy}}^{110} \quad \text{(3}^N\text{ terms)} \)

Note: \( R(T_{\text{easy}}^{\otimes N}) = (q + 1)^{N + o(N)} \) would imply \( \omega = 2 \)

Coppersmith and Winograd showed how to select \( \approx \left( \frac{3}{2^{2/3}} \right)^N \) terms that do not share variables (i.e., form a direct sum)

by zeroing variables (i.e., without increasing the rank)
The first CW construction: Analysis

Theorem [Coppermith and Winograd 87]

The tensor $T_{\text{easy}}^\otimes N$ can be converted into a direct sum of

$$\exp \left( \left( H \left( \frac{1}{3}, \frac{2}{3} \right) - o(1) \right) N \right) = \left( \frac{3}{2^{2/3}} \right)^{(1-o(1))N}$$

by zeroing variables (i.e., without increasing the rank)

$N/3$ copies of $T_{\text{easy}}^{011}$, $N/3$ copies of $T_{\text{easy}}^{101}$ and $N/3$ copies of $T_{\text{easy}}^{110}$.

Consider $T_{\text{easy}}^\otimes N = T_{\text{easy}}^{011} \otimes \cdots \otimes T_{\text{easy}}^{011} + \cdots + T_{\text{easy}}^{110} \otimes \cdots \otimes T_{\text{easy}}^{110}$ (3$^N$ terms)
Theorem [Coppermith and Winograd 87]

The tensor $T_{\text{easy}}^\otimes N$ can be converted into a direct sum of

$$\exp \left( \left( H \left( \frac{1}{3}, \frac{2}{3} \right) - o(1) \right) N \right) = \left( \frac{3}{2^{2/3}} \right)^{(1-o(1))N}$$

terms, each containing $\frac{N}{3}$ copies of $T_{\text{easy}}^{011}$, $\frac{N}{3}$ copies of $T_{\text{easy}}^{101}$ and $\frac{N}{3}$ copies of $T_{\text{easy}}^{110}$.

isomorphic to $[T_{\text{easy}}^{011}]^\otimes N/3 \otimes [T_{\text{easy}}^{101}]^\otimes N/3 \otimes [T_{\text{easy}}^{110}]^\otimes N/3 \cong \langle q^{N/3}, q^{N/3}, q^{N/3} \rangle$

Theorem (the asymptotic sum inequality, general form) [Schönhage 1981]

$$R \left( \bigoplus_{i=1}^{k} \langle m_i, n_i, p_i \rangle \right) \leq t \implies \sum_{i=1}^{k} (m_i n_i p_i)^{\omega/3} \leq t$$

Consequence: $\left( \frac{3}{2^{2/3}} \right)^{(1-o(1))N} \times q^{N\omega/3} \leq R(T_{\text{easy}}^\otimes N) \leq R(T_{\text{easy}})^N = (q+2)^N$

$$\implies \frac{3}{2^{2/3}} \times q^{\omega/3} \leq q + 2 \implies \omega \leq 2.403... \text{ for } q = 8$$
Idea behind the proof

\[ T_{\text{easy}} = T_{\text{easy}}^{011} + T_{\text{easy}}^{101} + T_{\text{easy}}^{110} \]

\[ T_{\text{easy}}^{011} = \sum_{i=1}^{q} x_0 \otimes y_i \otimes z_i \]

\[ T_{\text{easy}}^{101} = \sum_{i=1}^{q} x_i \otimes y_0 \otimes z_i \]

\[ T_{\text{easy}}^{110} = \sum_{i=1}^{q} x_i \otimes y_i \otimes z_0 \]

Consider \( N = 2 \)

\[ T_{\text{easy}}^\otimes^2 = (T_{\text{easy}}^{011} + T_{\text{easy}}^{101} + T_{\text{easy}}^{110}) \otimes (T_{\text{easy}}^{011} + T_{\text{easy}}^{101} + T_{\text{easy}}^{110}) \]

\[ = T_{\text{easy}}^{011} \otimes T_{\text{easy}}^{011} + T_{\text{easy}}^{011} \otimes T_{\text{easy}}^{101} + \cdots + T_{\text{easy}}^{110} \otimes T_{\text{easy}}^{110} \]

(9 terms)

\[ T_{\text{easy}}^{011} \otimes T_{\text{easy}}^{101} = \sum_{i,i'=0}^{q} \begin{pmatrix} 110110 & 101101 & 111001 \\ 011100 & 100111 & 110011 \end{pmatrix} \]

Tensor over \( (U_0 \otimes U_1) \otimes (V_1 \otimes V_0) \otimes (W_1 \otimes W_1) \)
\[ T_{\text{eas}y}^{011} \otimes T_{\text{eas}y}^{011} = \sum_{i,i'=0}^{q} (x_0 \otimes x_0) \otimes (y_i \otimes y_i') \otimes (z_i \otimes z_i') \]
tensor over \((U_0 \otimes U_0) \otimes (V_1 \otimes V_1) \otimes (W_1 \otimes W_1)\)

Idea behind the proof

\[
\begin{align*}
T_{\text{eas}y}^{011} \otimes T_{\text{eas}y}^{011} &= (T_{\text{eas}y}^{011} + T_{\text{eas}y}^{101} + T_{\text{eas}y}^{110}) \otimes (T_{\text{eas}y}^{011} + T_{\text{eas}y}^{101} + T_{\text{eas}y}^{110}) \\
&= T_{\text{eas}y}^{001111} \otimes T_{\text{eas}y}^{001111} + T_{\text{eas}y}^{011011} \otimes T_{\text{eas}y}^{011011} + \cdots + T_{\text{eas}y}^{111100} \otimes T_{\text{eas}y}^{111100} \quad \text{(9 terms)}
\end{align*}
\]

Consider \(N = 2\)

\[ T_{\text{eas}y}^{\otimes 2} = (T_{\text{eas}y}^{011} + T_{\text{eas}y}^{101} + T_{\text{eas}y}^{110}) \otimes (T_{\text{eas}y}^{011} + T_{\text{eas}y}^{101} + T_{\text{eas}y}^{110}) \]

remove this term
(e.g., by setting all variables in \(V_1 \otimes V_1\) to zero)

note: this removes more than one term!

SHARE VARIABLES

by setting all variables in \(U_1 \otimes U_0, V_0 \otimes V_0, W_1 \otimes W_1\) to zero

(label \(011011\))
Idea behind the proof

Conclusion: we can convert $T_{\text{easy}}^\otimes 2$ (a sum of 9 terms) into a direct sum of 2 terms

NEXT STEP

Consider $T_{\text{easy}}^\otimes N = T_{\text{easy}}^{011} \otimes \cdots \otimes T_{\text{easy}}^{011} + \cdots + T_{\text{easy}}^{110} \otimes \cdots \otimes T_{\text{easy}}^{110}$ ($3^N$ terms)

labels: $\begin{array}{c}
\bullet \ 0 \ 1 \ 1 \ \cdots \ 1 \\
\downarrow \ 0 \ \cdots \ 0 \ \\
3N
\end{array}$ \hspace{1cm} $\begin{array}{c}
\bullet \ 1 \ 1 \ 0 \ \cdots \ 0 \\
\\downarrow \ 1 \ \cdots \ 1 \\
3N
\end{array}$
Idea behind the proof

Theorem [Coppermith and Winograd 87]

The tensor $T_{\text{easy}}^\otimes N$ can be converted into a direct sum of

$$\exp \left( \left( H \left( \frac{1}{3}, \frac{2}{3} \right) - o(1) \right) N \right) = \left( \frac{3}{2^{2/3}} \right)^{(1-o(1))N}$$

terms, each containing $\frac{N}{3}$ copies of $T_{\text{easy}}^{011}$, $\frac{N}{3}$ copies of $T_{\text{easy}}^{101}$ and $\frac{N}{3}$ copies of $T_{\text{easy}}^{110}$.

We can obtain $\left( \frac{3}{2^{2/3}} \right)^{(1-o(1))N}$ labels of the form

$$\begin{align*}
\text{number of possibilities} & \approx \exp \left( H \left( \frac{1}{3}, \frac{2}{3} \right) N \right) \\
\end{align*}$$

that do not share a blue part, a red part or a green part.

The proof of this theorem is based on a complicated construction using the existence of dense sets of integers with no three-term arithmetic progression.
General Formulation of the Laser Method and Reinterpretation
The laser method: general formulation

For any tensor $T$, any $N \geq 1$ and any $\rho \in [2, 3]$ define $V_{\rho,N}(T)$ as the maximum of $\sum_{i=1}^{k} (m_i n_i p_i)^{\rho/3}$ over all restrictions of $T \otimes N$ isomorphic to $\bigoplus_{i=1}^{k} \langle m_i, n_i, p_i \rangle$

$$V_{\rho}(T) = \lim_{N \to \infty} V_{\rho,N}(T)^{1/N}$$

The value of $T$

This is the definition for symmetric tensors. Otherwise we use $V_{\rho}(T) = V_{\rho}(T \otimes \pi T \otimes \pi^2 T)^{1/3}$

$$V_{\rho}(\langle m, n, p \rangle) = (mnp)^{\rho/3}$$

This is an increasing function of $\rho$

$$V_{\rho}(T \oplus T') \geq V_{\rho}(T) + V_{\rho}(T') \quad V_{\rho}(T \otimes T') \geq V_{\rho}(T) \times V_{\rho}(T')$$

Theorem (the asymptotic sum inequality, general form) [Schönhage 1981]

$$\sum_{i=1}^{k} (m_i n_i p_i)^{\omega/3} \leq R \left( \bigoplus_{i=1}^{k} \langle m_i, n_i, p_i \rangle \right)$$
For any tensor $T$, any $N \geq 1$ and any $\rho \in [2, 3]$ define $V_{\rho,N}(T)$ as the maximum of $\sum_{i=1}^{k} (m_i n_i p_i)^{\rho/3}$ over all restrictions of $T \otimes N$ isomorphic to $\bigoplus_{i=1}^{k} \langle m_i, n_i, p_i \rangle$.

$V_{\rho}(T) = \lim_{N \to \infty} V_{\rho,N}(T)^{1/N}$

The value of $T$

**Theorem [Coppermmith and Winograd 87]**

The tensor $T_{\text{easy}} \otimes N$ can be converted into a direct sum of

$$\exp \left( \left( H \left( \frac{1}{3}, \frac{2}{3} \right) - o(1) \right) N \right) = \left( \frac{3}{2^{2/3}} \right)^{(1-o(1))N}$$

terms, each containing $\frac{N}{3}$ copies of $T_{\text{easy}}^{011}$, $\frac{N}{3}$ copies of $T_{\text{easy}}^{101}$ and $\frac{N}{3}$ copies of $T_{\text{easy}}^{110}$.

isomorphic to $[T_{\text{easy}}^{011}] \otimes N/3 \otimes [T_{\text{easy}}^{101}] \otimes N/3 \otimes [T_{\text{easy}}^{110}] \otimes N/3 \cong \langle q^{N/3}, q^{N/3}, q^{N/3} \rangle$

$$V_{\rho,N}(T_{\text{easy}}) \geq \left( \frac{3}{2^{2/3}} \right)^{(1-o(1))N} \times q^{\rho N/3} \quad \Rightarrow \quad V_{\rho}(T_{\text{easy}}) \geq \frac{3}{2^{2/3}} \times q^{\rho/3}$$
The laser method: general formulation

For any tensor $T$, any $N \geq 1$ and any $\rho \in [2, 3]$ define $V_{\rho,N}(T)$ as the maximum of $\sum_{i=1}^{k} (m_i n_i p_i)^{\rho/3}$ over all restrictions of $T \otimes T$ isomorphic to $\bigoplus_{i=1}^{k} \langle m_i, n_i, p_i \rangle$.

$$V_{\rho}(T) = \lim_{N \to \infty} V_{\rho,N}(T)^{1/N}$$

The value of $T$ for instance, $V_{\omega}(\langle m, n, p \rangle) = (mnp)^{\rho}$

Theorem (the asymptotic sum inequality, general form) [Schönhage 1981]

$$\sum_{i=1}^{k} (m_i n_i p_i)^{\omega/3} \leq R \left( \bigoplus_{i=1}^{k} \langle m_i, n_i, p_i \rangle \right)$$

Theorem (simple generalization of the asymptotic sum inequality)

$$V_{\omega}(T) \leq R(T)$$
The laser method: general formulation

Consider three vector spaces $U$, $V$ and $W$ over $\mathbb{F}$.

A tensor $T$ over $(U, V, W)$ is an element of $U \otimes V \otimes W$.

Assume that $U$, $V$ and $W$ are decomposed as

$$U = \bigoplus_{i \in I} U_i \quad V = \bigoplus_{j \in J} V_j \quad W = \bigoplus_{k \in K} W_k$$

for some $I, J, K \subseteq \mathbb{Z}$.

The tensor $T$ is a partitioned tensor (with respect to this decomposition) if it can be written as

$$T = \sum_{(i,j,k) \in I \times J \times K} T_{ijk}$$

where $T_{ijk} \in U_i \otimes V_j \otimes W_k$ for each $(i, j, k) \in I \times J \times K$.

The support of the tensor:

$$\text{supp}(T) = \{(i, j, k) \in I \times J \times K \mid T_{ijk} \neq 0\}$$

each non-zero $T_{ijk}$ is called a component of $T$.

We say that the tensor is **tight** if there exists some integer $d$ such that $i + j + k = d$ for all $(i, j, k) \in \text{supp}(T)$.
Example: The first CW construction

\[ U = U_0 \oplus U_1, \quad \text{where } U_0 = \text{span}\{x_0\} \text{ and } U_1 = \text{span}\{x_1, \ldots, x_q\} \]
\[ V = V_0 \oplus V_1, \quad \text{where } V_0 = \text{span}\{y_0\} \text{ and } V_1 = \text{span}\{y_1, \ldots, y_q\} \]
\[ W = W_0 \oplus W_1, \quad \text{where } W_0 = \text{span}\{z_0\} \text{ and } W_1 = \text{span}\{z_1, \ldots, z_q\} \]

\[ T_{\text{easy}} = T_{\text{easy}}^{011} + T_{\text{easy}}^{101} + T_{\text{easy}}^{110}, \]

where
\[ T_{\text{easy}}^{011} = \sum_{i=1}^{q} x_0 \otimes y_i \otimes z_i \]
\[ T_{\text{easy}}^{101} = \sum_{i=1}^{q} x_i \otimes y_0 \otimes z_i \]
\[ T_{\text{easy}}^{110} = \sum_{i=1}^{q} x_i \otimes y_i \otimes z_0 \]

\[ \text{tensor over } (U_0, V_1, W_1) \]
\[ \text{tensor over } (U_1, V_0, W_1) \]
\[ \text{tensor over } (U_1, V_1, W_0) \]

\[ \text{supp}(T_{\text{easy}}) = \{(0, 1, 1), (1, 0, 1), (1, 1, 0)\} \]

it is tight, since \( i + j + k = 2 \) for all \((i, j, k) \in \text{supp}(T_{\text{easy}})\)
The laser method: general formulation

Main Theorem [LG 14] (reinterpretation of prior works)

For any tight partitioned tensor $T$, any probability distribution $P$ over $\text{supp}(T)$, and any $\rho \in [2, 3]$, we have

$$\log(V_\rho(T)) \geq \sum_{\ell=1}^{3} \frac{H(P_\ell)}{3} + \sum_{(i,j,k) \in \text{supp}(T)} P(i,j,k) \log(V_\rho(T_{ijk})) - \Gamma(P).$$

$H$: entropy
$P_\ell$: projection of $P$ along the $\ell$-th coordinate (= marginal distribution)
$\Gamma(P)$: to be defined later (zero in the case of simple tensors)

Conclusion: we can compute a lower bound on the value of $T$ if we know a lower bound on the value of each component, we can then obtain an upper bound on $\omega$ via $V_\omega(T) \leq R(T)$.

Concretely, we use $V_\rho(T) \geq R(T) \implies \omega \leq \rho$ and do a binary search on $\rho$. 
Example: The first CW construction

**Main Theorem [LG 14] (reinterpretation of prior works)**

For any tight partitioned tensor $T$, any probability distribution $P$ over $\text{supp}(T)$, and any $\rho \in [2, 3]$, we have

$$\log(V_\rho(T)) \geq \sum_{\ell=1}^{3} \frac{H(P_\ell)}{3} + \sum_{(i,j,k) \in \text{supp}(T)} P(i, j, k) \log(V_\rho(T_{ijk})) - \Gamma(P).$$

$H$: entropy

$P_\ell$: projection of $P$ along the $\ell$-th coordinate (= marginal distribution)

$\Gamma(P)$: to be defined later (zero in the case of simple tensors)

$\text{supp}(T_{\text{easy}}) = \{(0, 1, 1), (1, 0, 1), (1, 1, 0)\}$

$P(0,1,1) = P(1,0,1) = P(1,1,0) = 1/3$

$\Gamma(P) = 0$

$P_1(0) = 1/3$, $P_1(1) = 2/3$ and $P_2 = P_3 = P_1$

$V_\rho(T_{\text{easy}}^{011}) = V_\rho(\langle 1, 1, q \rangle) = q^{\rho/3}$

$V_\rho(T_{\text{easy}}^{101}) = V_\rho(\langle q, 1, 1 \rangle) = q^{\rho/3}$

$V_\rho(T_{\text{easy}}^{110}) = V_\rho(\langle 1, q, 1 \rangle) = q^{\rho/3}$

$$\log(V_\rho(T_{\text{easy}})) \geq H\left(\frac{1}{3}, \frac{2}{3}\right) + \frac{1}{3} \log\left(q^{\rho/3}\right) + \frac{1}{3} \log\left(q^{\rho/3}\right) + \frac{1}{3} \log\left(q^{\rho/3}\right)$$
The tensor $T_{\text{easy}}^\otimes N$ can be converted into a direct sum of terms, each containing $\frac{N}{3}$ copies of $T_{\text{easy}}^{011}$, $\frac{N}{3}$ copies of $T_{\text{easy}}^{101}$, and $\frac{N}{3}$ copies of $T_{\text{easy}}^{110}$.

$$V_{\rho, N}(T_{\text{easy}}) \geq \left( \frac{3}{2^{2/3}} \right)^{(1-o(1))N} \times q^{\rho N/3}$$

$$V_{\rho}(T_{\text{easy}}) \geq \frac{3}{2^{2/3}} \times q^{\rho/3}$$
The laser method: general formulation

Main Theorem [LG 14]

For any tight partitioned tensor $T$, any probability distribution $P$ over $\text{supp}(T)$, and any $\rho \in [2, 3]$, we have

$$\log(V_\rho(T)) \geq \sum_{\ell=1}^{3} \frac{H(P_\ell)}{3} + \sum_{(i,j,k) \in \text{supp}(T)} P(i, j, k) \log(V_\rho(T_{ijk})) - \Gamma(P).$$

Interpretation: the laser method enables us to convert (by zeroing variables)

$$T^\otimes N$$

into a direct sum of

$$\exp\left(\left(\sum_{\ell=1}^{3} \frac{H(P_\ell)}{3} - \Gamma(P) - o(1)\right)N\right)$$

terms, each isomorphic to

$$\bigotimes_{(i,j,k) \in \text{supp}(T)} \left[T_{ijk}\right] \otimes P(i,j,k) N$$
The second CW construction

Let $q$ be a positive integer.

Consider three vector spaces $U$, $V$ and $W$ of dimension $q + 2$ over $\mathbb{F}$.

$$U = \text{span}\{x_0, \ldots, x_q, x_{q+1}\} \quad W = \text{span}\{z_0, \ldots, z_q, z_{q+1}\}$$

$$V = \text{span}\{y_0, \ldots, y_q, y_{q+1}\}$$

Coppersmith and Winograd (1987) considered the following tensor:

$$T_{CW} = T_{easy} + T_{CW}^{002} + T_{CW}^{020} + T_{CW}^{200}$$

$$R(T_{CW}) = q + 2$$

$$T_{CW} = T_{CW}^{011} + T_{CW}^{101} + T_{CW}^{110} + T_{CW}^{002} + T_{CW}^{020} + T_{CW}^{200},$$

where

\begin{align*}
T_{CW}^{011} &= T_{easy}^{011} \\
T_{CW}^{101} &= T_{easy}^{101} \\
T_{CW}^{110} &= T_{easy}^{110} \\
T_{CW}^{002} &= x_0 \otimes y_0 \otimes z_{q+1} \cong \langle 1, 1, 1 \rangle \\
T_{CW}^{020} &= x_0 \otimes y_{q+1} \otimes z_0 \cong \langle 1, 1, 1 \rangle \\
T_{CW}^{200} &= x_{q+1} \otimes y_0 \otimes z_0 \cong \langle 1, 1, 1 \rangle.
\end{align*}
The second CW construction

\[ U = \text{span}\{x_0, \ldots, x_q, x_{q+1}\} \quad W = \text{span}\{z_0, \ldots, z_q, z_{q+1}\} \]

\[ V = \text{span}\{y_0, \ldots, y_q, y_{q+1}\} \]

\[ U = U_0 \oplus U_1 \oplus U_2, \quad \text{where } U_0 = \text{span}\{x_0\}, U_1 = \text{span}\{x_1, \ldots, x_q\} \text{ and } U_2 = \text{span}\{x_{q+1}\} \]

\[ V = V_0 \oplus V_1 \oplus V_2, \quad \text{where } V_0 = \text{span}\{y_0\}, V_1 = \text{span}\{y_1, \ldots, y_q\} \text{ and } V_2 = \text{span}\{y_{q+1}\} \]

\[ W = W_0 \oplus W_1 \oplus W_2, \quad \text{where } W_0 = \text{span}\{z_0\}, W_1 = \text{span}\{z_1, \ldots, z_q\} \text{ and } W_2 = \text{span}\{z_{q+1}\} \]

\[ T_{\text{CW}} = T_{\text{CW}}^{011} + T_{\text{CW}}^{101} + T_{\text{CW}}^{110} + T_{\text{CW}}^{002} + T_{\text{CW}}^{020} + T_{\text{CW}}^{200} \]

This is not a direct sum

\[ T_{\text{CW}}^{011} \quad \text{tensor over } (U_0, V_1, W_1) \]

\[ T_{\text{CW}}^{101} \quad \text{tensor over } (U_1, V_0, W_1) \]

\[ T_{\text{CW}}^{110} \quad \text{tensor over } (U_1, V_1, W_0) \]

\[ T_{\text{CW}}^{002} = x_0 \otimes y_0 \otimes z_{q+1} \cong \langle 1, 1, 1 \rangle \quad \text{tensor over } (U_0, V_0, W_2) \]

\[ T_{\text{CW}}^{020} = x_0 \otimes y_{q+1} \otimes z_0 \cong \langle 1, 1, 1 \rangle \quad \text{tensor over } (U_0, V_2, W_0) \]

\[ T_{\text{CW}}^{200} = x_{q+1} \otimes y_0 \otimes z_0 \cong \langle 1, 1, 1 \rangle \quad \text{tensor over } (U_2, V_0, W_0) \]
The second CW construction: laser method

\[ \supp(T_{CW}) = \{(0, 1, 1), (1, 0, 1), (1, 1, 0), (0, 0, 2), (0, 2, 0), (2, 0, 0)\} \]

\[ V_\rho(T_{CW}^{002}) = V_\rho(T_{CW}^{020}) = V_\rho(T_{CW}^{200}) = 1 \quad V_\rho(T_{CW}^{011}) = V_\rho(T_{CW}^{101}) = V_\rho(T_{CW}^{110}) = q^{\rho/3} \]

Take

\[ P(0, 1, 1) = P(1, 0, 1) = P(1, 1, 0) = \alpha \]
\[ P(0, 0, 2) = P(0, 2, 0) = P(2, 0, 0) = (1/3 - \alpha) \]
\[ P_1(0) = \alpha + 2(1/3 - \alpha), \quad P_1(1) = 2\alpha, \quad P_1(2) = (1/3 - \alpha) \]

\[ T_{CW} = T_{CW}^{011} + T_{CW}^{101} + T_{CW}^{110} + T_{CW}^{002} + T_{CW}^{020} + T_{CW}^{200} \]

\[ T_{CW}^{011} \text{ tensor over } (U_0, V_1, W_1) \]
\[ T_{CW}^{101} \text{ tensor over } (U_1, V_0, W_1) \]
\[ T_{CW}^{110} \text{ tensor over } (U_1, V_1, W_0) \]
\[ T_{CW}^{002} = x_0 \otimes y_0 \otimes z_{q+1} \cong \langle 1, 1, 1 \rangle \text{ tensor over } (U_0, V_0, W_2) \]
\[ T_{CW}^{020} = x_0 \otimes y_{q+1} \otimes z_0 \cong \langle 1, 1, 1 \rangle \text{ tensor over } (U_0, V_2, W_0) \]
\[ T_{CW}^{200} = x_{q+1} \otimes y_0 \otimes z_0 \cong \langle 1, 1, 1 \rangle \text{ tensor over } (U_2, V_0, W_0) \]
**The second CW construction: laser method**

\[ \text{supp}(T_{CW}) = \{(0, 1, 1), (1, 0, 1), (1, 1, 0), (0, 0, 2), (0, 2, 0), (2, 0, 0)\} \]

\[ V_\rho(T_{CW}^{002}) = V_\rho(T_{CW}^{020}) = V_\rho(T_{CW}^{200}) = 1 \quad V_\rho(T_{CW}^{011}) = V_\rho(T_{CW}^{101}) = V_\rho(T_{CW}^{110}) = q^{\rho/3} \]

take

\[ P(0, 1, 1) = P(1, 0, 1) = P(1, 1, 0) = \alpha \]
\[ P(0, 0, 2) = P(0, 2, 0) = P(2, 0, 0) = (1/3 - \alpha) \]
\[ P_1(0) = \alpha + 2(1/3 - \alpha), \quad P_1(1) = 2\alpha, \quad P_1(2) = (1/3 - \alpha) \]

**Main Theorem [LG 14]**

For any tight partitioned tensor \( T \), any probability distribution \( P \) over \( \text{supp}(T) \), and any \( \rho \in [2, 3] \), we have

\[
\log(V_\rho(T)) \geq \sum_{\ell=1}^{3} \frac{H(P_\ell)}{3} + \sum_{(i,j,k) \in \text{supp}(T)} P(i, j, k) \log(V_\rho(T_{ijk})) - \Gamma(P).
\]

\[
\implies \log(V_\rho(T_{CW})) \geq H \left( \frac{2}{3} - \alpha, 2\alpha, \frac{1}{3} - \alpha \right) + \log(q^{\alpha \omega})
\]

combined with \( V_\omega(T_{CW}) \leq R(T_{CW}) = q + 2 \)

this gives \( \omega \leq 2.38718... \) for \( q = 6 \) and \( \alpha = 0.3173 \)
## Analysis of the second construction

Analysis of the $m$-th power of the tensor by CW

<table>
<thead>
<tr>
<th>$m$</th>
<th>Upper bound</th>
<th>Number of variables in the optimization problem</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\omega &lt; 2.3871900$</td>
<td>1</td>
<td>CW (1987)</td>
</tr>
<tr>
<td>2</td>
<td>$\omega &lt; 2.3754770$</td>
<td>3</td>
<td>CW (1987)</td>
</tr>
<tr>
<td>4</td>
<td>$\omega &lt; 2.3729269$</td>
<td>9</td>
<td>Stothers (2010)</td>
</tr>
<tr>
<td>8</td>
<td>$\omega &lt; 2.3729$</td>
<td>29</td>
<td>Vassilevska Williams (2012)</td>
</tr>
<tr>
<td>16</td>
<td>$\omega &lt; 2.3728640$</td>
<td>101</td>
<td>Le Gall (2014)</td>
</tr>
<tr>
<td>32</td>
<td>$\omega &lt; 2.3728639$</td>
<td>373</td>
<td>Le Gall (2014)</td>
</tr>
</tbody>
</table>
Analysis of the second power

\[ T_{CW}^{\otimes 2} = (T_{CW}^{011} + T_{CW}^{101} + T_{CW}^{110} + T_{CW}^{002} + T_{CW}^{020} + T_{CW}^{200}) \otimes 2 \] (36 terms)

\[ R(T_{CW}^{\otimes 2}) \leq (q + 2)^2 \]

Idea: rewrite it as a (non-direct) sum of 15 terms by regrouping terms

\[ T_{CW}^{\otimes 2} = T^{400} + T^{040} + T^{004} + T^{310} + T^{301} + T^{103} + T^{130} + T^{013} \]
\[ + T^{031} + T^{220} + T^{202} + T^{022} + T^{211} + T^{121} + T^{112}, \]

where

\[ T^{400} = T_{CW}^{200} \otimes T_{CW}^{200}, \]
\[ T^{310} = T_{CW}^{200} \otimes T_{CW}^{110} + T_{CW}^{110} \otimes T_{CW}^{200}, \]
\[ T^{220} = T_{CW}^{200} \otimes T_{CW}^{020} + T_{CW}^{020} \otimes T_{CW}^{200} + T_{CW}^{110} \otimes T_{CW}^{110}, \]
\[ T^{211} = T_{CW}^{200} \otimes T_{CW}^{011} + T_{CW}^{011} \otimes T_{CW}^{200} + T_{CW}^{110} \otimes T_{CW}^{101} + T_{CW}^{101} \otimes T_{CW}^{110}, \]

and the other 11 terms are obtained by permuting the variables (e.g., \( T^{040} = T_{CW}^{020} \otimes T_{CW}^{020} \)).
Analysis of the second power

\[ \text{supp}(T_{CW}^\otimes 2) = \{(4, 0, 0), \ldots, (0, 0, 4), (3, 1, 0), \ldots, (0, 1, 3), (2, 2, 0), \ldots, (0, 2, 2), (2, 1, 1), \ldots, (1, 1, 2)\} \]

lower bounds on the values of each component can be computed (recursively)

choice of distribution: \( P(4, 0, 0) = \ldots = P(0, 0, 4) = \alpha, \quad P(3, 1, 0) = \ldots = P(0, 1, 3) = \beta \)

\( P(2, 2, 0) = \ldots = P(0, 2, 2) = \gamma, \quad P(2, 1, 1) = \ldots = P(1, 1, 2) = \delta \)

\[ T_{CW}^\otimes 2 = T^{400} + T^{040} + T^{004} + T^{310} + T^{301} + T^{103} + T^{130} + T^{013} \]

\[ + T^{031} + T^{220} + T^{202} + T^{022} + T^{211} + T^{121} + T^{112}, \]

where

\[ T^{400} = T_{CW}^{200} \otimes T_{CW}^{200}, \]

\[ T^{310} = T_{CW}^{200} \otimes T_{CW}^{110} + T_{CW}^{110} \otimes T_{CW}^{200}, \]

\[ T^{220} = T_{CW}^{200} \otimes T_{CW}^{020} + T_{CW}^{020} \otimes T_{CW}^{200} + T_{CW}^{110} \otimes T_{CW}^{110}, \]

\[ T^{211} = T_{CW}^{200} \otimes T_{CW}^{011} + T_{CW}^{011} \otimes T_{CW}^{200} + T_{CW}^{110} \otimes T_{CW}^{101} + T_{CW}^{101} \otimes T_{CW}^{110}, \]

and the other 11 terms are obtained by permuting the variables (e.g., \( T^{040} = T_{CW}^{020} \otimes T_{CW}^{020} \)).
Analysis of the second power

\[ \text{supp}(T_{\text{CW}}^\otimes 2) = \{(4, 0, 0), \ldots, (0, 0, 4), (3, 1, 0), \ldots, (0, 1, 3), (2, 2, 0), \ldots, (0, 2, 2), (2, 1, 1), \ldots, (1, 1, 2)\} \]

3 permutations 6 permutations 3 permutations 3 permutations

lower bounds on the values of each component can be computed (recursively)

choice of distribution: \( P(4, 0, 0) = \ldots = P(0, 0, 4) = \alpha, \ P(3, 1, 0) = \ldots = P(0, 1, 3) = \beta \)
\( P(2, 2, 0) = \ldots = P(0, 2, 2) = \gamma, \ P(2, 1, 1) = \ldots = P(1, 1, 2) = \delta \)

we have \( \Gamma(P) = 0 \)
Analysis of the second power

\[ \text{supp}(T_{\text{CW}}^{\otimes 2}) = \{(4, 0, 0), \ldots, (0, 0, 4), (3, 1, 0), \ldots, (0, 1, 3), (2, 2, 0), \ldots, (0, 2, 2), (2, 1, 1), \ldots, (1, 1, 2)\} \]

lower bounds on the values of each component can be computed (recursively)

choice of distribution:

\[ P(4, 0, 0) = \ldots = P(0, 0, 4) = \alpha, \quad P(3, 1, 0) = \ldots = P(0, 1, 3) = \beta \]
\[ P(2, 2, 0) = \ldots = P(0, 2, 2) = \gamma, \quad P(2, 1, 1) = \ldots = P(1, 1, 2) = \delta \]

we have \( \Gamma(P) = 0 \)

Main Theorem [LG 14]

For any tight partitioned tensor \( T \), any probability distribution \( P \) over \( \text{supp}(T) \), and any \( \rho \in [2, 3] \), we have

\[
\log(V_{\rho}(T)) \geq \sum_{\ell=1}^{3} \frac{H(P_{\ell})}{3} + \sum_{(i,j,k) \in \text{supp}(T)} P(i, j, k) \log(V_{\rho}(T_{ijk})) - \Gamma(P).
\]

Theorem

\[ V_{\omega}(T) \leq R(T) \quad \Rightarrow \omega \leq 2.3755 \ldots \text{ for } q = 6 \text{ and } \alpha = 0.00023, \beta = 0.0125, \gamma = 0.10254 \text{ and } \delta = 0.2056 \]
### Analysis of the second power

<table>
<thead>
<tr>
<th>$m$</th>
<th>Upper bound</th>
<th>Number of variables in the optimization problem</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\omega &lt; 2.3871900$</td>
<td>1</td>
<td>CW (1987)</td>
</tr>
<tr>
<td>2</td>
<td>$\omega &lt; 2.3754770$</td>
<td>3</td>
<td>CW (1987)</td>
</tr>
<tr>
<td>4</td>
<td>$\omega &lt; 2.3729269$</td>
<td>9</td>
<td>Stothers (2010)</td>
</tr>
<tr>
<td>8</td>
<td>$\omega &lt; 2.3729$</td>
<td>29</td>
<td>Vassilevska Williams (2012)</td>
</tr>
<tr>
<td>16</td>
<td>$\omega &lt; 2.3728640$</td>
<td>101</td>
<td>Le Gall (2014)</td>
</tr>
<tr>
<td>32</td>
<td>$\omega &lt; 2.3728639$</td>
<td>373</td>
<td>Le Gall (2014)</td>
</tr>
</tbody>
</table>

What about the third power (using similar merging schemes)?

→ this does not give any improvement
Analysis of the fourth power

\[ T_{CW}^{\otimes 4} = (T_{CW}^{011} + T_{CW}^{101} + T_{CW}^{110} + T_{CW}^{002} + T_{CW}^{020} + T_{CW}^{200})^{\otimes 4} \] (6^4 terms)

\[ R(T_{CW}^{\otimes 4}) \leq (q + 2)^4 \]

Idea: rewrite it as a (non-direct) sum of a smaller number of terms by regrouping terms

\[ T_{CW}^{\otimes 4} = T^{800} + T^{710} + T^{620} + T^{611} + T^{530} + T^{521} + T^{440} + T^{431} + T^{422} + T^{332} \] + permutations of these terms

\[ T^{080}, T^{008}, T^{701}, T^{107}, T^{170}, T^{017}, T^{071}, \ldots \]

10-1=9 parameters for the probability distribution

this time \( \Gamma(P) \neq 0 \)
The laser method: general formulation

Main Theorem [LG 14]

For any tight partitioned tensor $T$, any probability distribution $P$ over $\text{supp}(T)$, and any $\rho \in [2, 3]$, we have

$$\log(V_\rho(T)) \geq \sum_{\ell=1}^{3} \frac{H(P_\ell)}{3} + \sum_{(i,j,k) \in \text{supp}(T)} P(i, j, k) \log(V_\rho(T_{ijk})) - \Gamma(P).$$

$H$: entropy

$P_\ell$: projection of $P$ along the $\ell$-th coordinate (= marginal distribution)

$\Gamma(P)$: to be defined later (zero in the case of simple tensors)
The laser method: general formulation

**Main Theorem [LG 14]**

For any tight partitioned tensor $T$, any probability distribution $P$ over $\text{supp}(T)$, and any $\rho \in [2, 3]$, we have

$$\log(V_{\rho}(T)) \geq \sum_{\ell=1}^{3} \frac{H(P_{\ell})}{3} + \sum_{(i,j,k) \in \text{supp}(T)} P(i, j, k) \log(V_{\rho}(T_{ijk})) - \Gamma(P).$$

$H$: entropy  
$P_{\ell}$: projection of $P$ along the $\ell$-th coordinate (= marginal distribution)  
$\Gamma(P)$: to be defined later (zero in the case of simple tensors)

\[ \Gamma(P) = \max[H(Q)] - H(P) \]

where the max is over all distributions $Q$ over $\text{supp}(T)$ such that $P_1 = Q_1$, $P_2 = Q_2$ and $P_3 = Q_3$.

when the structure of support is simple, we typically have $P_1 = Q_1$, $P_2 = Q_2$, $P_3 = Q_3 \implies P = Q$ and thus $\Gamma(P) = 0$.
The laser method: general formulation

Interpretation: the laser method enables us to convert (by zeroing variables) $T^\otimes N$ into a direct sum of

$$
\exp \left( \sum_{\ell=1}^{3} \frac{H(P_\ell)}{3} - \Gamma(P) - o(1) \right) N
$$

terms, each isomorphic to $[T^{i,j,k}] \otimes P(i,j,k) N$ “type $P$”

we can control only the choice of the marginal distributions $P_1, P_2$ and $P_3$
what we obtain is a (non-direct) sum of all “type Q” terms
the most frequent terms are those with $Q$ maximizing $H(Q)$
the fact that “type $P$” are not the most frequent introduces the penalty term $-\Gamma(P)$

$$
\Gamma(P) = \max[H(Q)] - H(P)
$$
where the max is over all distributions $Q$ over $\text{supp}(T)$ such that $P_1 = Q_1, P_2 = Q_2$ and $P_3 = Q_3$
The laser method: computing the bound

**Main Theorem [LG 14]**

For any tight partitioned tensor $T$, any probability distribution $P$ over $\text{supp}(T)$, and any $\rho \in [2, 3]$, we have

$$\log(V_\rho(T)) \geq \sum_{\ell=1}^{3} \frac{H(P_\ell)}{3} + \sum_{(i,j,k) \in \text{supp}(T)} P(i,j,k) \log(V_\rho(T_{ijk})) - \Gamma(P).$$

How to find the best distribution for a given $\rho$?

- Assume that (a lower bound on) each $V_\rho(T_{ijk})$ is known.

If $\Gamma(P) = 0$ for all distributions $P$, the best distribution can be done efficiently (numerically) using convex optimization:

- Maximization of a **concave function** under **linear constraints**.
The laser method: computing the bound

Main Theorem [LG 14]

For any tight partitioned tensor $T$, any probability distribution $P$ over $\text{supp}(T)$, and any $\rho \in [2, 3]$, we have

\[
\log(V_\rho(T)) \geq \sum_{\ell=1}^{3} \frac{H(P_\ell)}{3} + \sum_{(i,j,k) \in \text{supp}(T)} P(i, j, k) \log(V_\rho(T_{ijk})) - \Gamma(P).
\]

How to find the best distribution for a given $\rho$?

Assume that (a lower bound on) each $V_\rho(T_{ijk})$ is known.

In general:

\[
\Gamma(P) = \max[H(Q)] - H(P)
\]

where the max is over all distributions $Q$ over $\text{supp}(T)$ such that $P_1 = Q_1$, $P_2 = Q_2$ and $P_3 = Q_3$

Hard to solve, but can be done up to the 4th power of the CW tensor [Stothers 10]
The laser method: computing the bound

- **Analysis of the $m$-th power of the tensor by CW**

<table>
<thead>
<tr>
<th>$m$</th>
<th>Upper bound</th>
<th>Number of variables in the optimization problem</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\omega &lt; 2.3871900$</td>
<td>1</td>
<td>CW (1987)</td>
</tr>
<tr>
<td>2</td>
<td>$\omega &lt; 2.3754770$</td>
<td>3</td>
<td>CW (1987)</td>
</tr>
<tr>
<td>4</td>
<td>$\omega &lt; 2.3729269$</td>
<td>9</td>
<td>Stothers (2010)</td>
</tr>
<tr>
<td>8</td>
<td>$\omega &lt; 2.3729$</td>
<td>29</td>
<td>Vassilevska Williams (2012)</td>
</tr>
<tr>
<td>16</td>
<td>$\omega &lt; 2.3728640$</td>
<td>101</td>
<td>Le Gall (2014)</td>
</tr>
<tr>
<td>32</td>
<td>$\omega &lt; 2.3728639$</td>
<td>373</td>
<td>Le Gall (2014)</td>
</tr>
</tbody>
</table>

In general:

$$\Gamma(P) = \max[H(Q)] - H(P)$$

where the max is over all distributions $Q$ over $\text{supp}(T)$

such that $P_1 = Q_1$, $P_2 = Q_2$ and $P_3 = Q_3$

hard to solve, but can be done up to the 4th power of the CW tensor [Stothers 10]
Main Theorem [LG 14]

For any tight partitioned tensor $T$, any probability distribution $P$ over supp$(T)$, and any $\rho \in [2, 3]$, we have

$$\log(V_\rho(T)) \geq \sum_{\ell=1}^{3} \frac{H(P_{\ell})}{3} + \sum_{(i,j,k) \in \text{supp}(T)} P(i,j,k) \log(V_\rho(T_{ijk})) - \Gamma(P).$$

How to find the best distribution for a given $\rho$?

- Assume that (a lower bound on) each $V_\rho(T_{ijk})$ is known.

In general:

$$\Gamma(P) = \max[H(Q)] - H(P)$$

where the max is over all distributions $Q$ over supp$(T)$ such that $P_1 = Q_1$, $P_2 = Q_2$ and $P_3 = Q_3$.

Hard to solve, but can be done up to the 4th power of the CW tensor [Stothers 10].

Simplification: restrict the search to the set of distributions $P$ such that $\Gamma(P) = 0$.

Still hard to solve, but can be done up to the 8th power of the CW tensor [Vassilevska-Williams 12].
The laser method: computing the bound

Simplification: restrict the search to the set of distributions $P$ such that $\Gamma(P) = 0$
still hard to solve, but can be done up to the 8th power of the CW tensor

<table>
<thead>
<tr>
<th>$m$</th>
<th>Upper bound</th>
<th>Number of variables in the optimization problem</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\omega &lt; 2.3871900$</td>
<td>1</td>
<td>CW (1987)</td>
</tr>
<tr>
<td>2</td>
<td>$\omega &lt; 2.3754770$</td>
<td>3</td>
<td>CW (1987)</td>
</tr>
<tr>
<td>4</td>
<td>$\omega &lt; 2.3729269$</td>
<td>9</td>
<td>Stothers (2010)</td>
</tr>
<tr>
<td>8</td>
<td>$\omega &lt; 2.3729$</td>
<td>29</td>
<td>Vassilevska Williams (2012)</td>
</tr>
<tr>
<td>16</td>
<td>$\omega &lt; 2.3728640$</td>
<td>101</td>
<td>Le Gall (2014)</td>
</tr>
<tr>
<td>32</td>
<td>$\omega &lt; 2.3728639$</td>
<td>373</td>
<td>Le Gall (2014)</td>
</tr>
</tbody>
</table>

analysis of the $m$-th power of the tensor by CW

[Vassilevska-Williams 12]
The laser method: computing the bound

Main Theorem [LG 14]

For any tight partitioned tensor $T$, any probability distribution $P$ over $\text{supp}(T)$, and any $\rho \in [2, 3]$, we have

$$\log(V_\rho(T)) \geq \frac{3}{\sum_{\ell=1}^{3} H(P_\ell)} + \sum_{(i,j,k) \in \text{supp}(T)} P(i,j,k) \log(V_\rho(T_{ijk})) - \Gamma(P),$$

where $H$ is linear and $V$ is concave.

How to find the best distribution for a given $\rho$?

Assume that (a lower bound on) each $V_\rho(T_{ijk})$ is known.

In general:

$$\Gamma(P) = \max[H(Q)] - H(P)$$

where the max is over all distributions $Q$ over $\text{supp}(T)$ such that $P_1 = Q_1$, $P_2 = Q_2$ and $P_3 = Q_3$.

Hard to solve, but can be done up to the 4th power of the CW tensor [Stothers 10]

Simplification: restrict the search to the set of distributions $P$ such that $\Gamma(P) = 0$.

Still hard to solve, but can be done up to the 8th power of the CW tensor [Vassilevska-Williams 12]
For any tight partitioned tensor $T$, any probability distribution $P$ over $\text{supp}(T)$, and any $\rho \in [2, 3]$, we have

$$\log(V_\rho(T)) \geq \sum_{\ell=1}^{3} \frac{H(P_\ell)}{3} + \sum_{(i,j,k) \in \text{supp}(T)} P(i, j, k) \log(V_\rho(T_{ijk})) - \Gamma(P).$$

\[ \Gamma(P) = \max[H(Q)] - H(P) \]

where the max is over all distributions $Q$ over $\text{supp}(T)$ such that $P_1 = Q_1$, $P_2 = Q_2$ and $P_3 = Q_3$.

Efficient method to find a solution [LG 14] (close to the optimal solution):
Main Theorem [LG 14]

For any tight partitioned tensor $T$, any probability distribution $P$ over $\text{supp}(T)$, and any $\rho \in [2, 3]$, we have

$$\log(V_\rho(T)) \geq \sum_{\ell=1}^{3} \frac{H(P_\ell)}{3} + \sum_{(i,j,k) \in \text{supp}(T)} P(i,j,k) \log(V_\rho(T_{ijk})) - \Gamma(P).$$

where the max is over all distributions $Q$ over $\text{supp}(T)$ such that $P_1 = Q_1$, $P_2 = Q_2$ and $P_3 = Q_3$

call this expression $f(P)$

Efficient method to find a solution [LG 14] (close to the optimal solution):

1. find a distribution $P$ that maximizes $f(P)$, and call it $\hat{P}$
   concave objective function, linear constraints
2. find the distribution $Q$ that maximizes $H(Q)$ under the constraints $Q_1 = \hat{P}_1$, $Q_2 = \hat{P}_2$ and $Q_3 = \hat{P}_3$. Call it $\hat{Q}$.
   concave objective function, linear constraints
3. output $f(\hat{Q})$

Since $\Gamma(\hat{Q}) = 0$, we have $\log(V_\rho(T)) \geq f(\hat{Q})$ from the theorem
## Analysis of power 16 and 32

An analysis of the $m$-th power of the tensor by CW

<table>
<thead>
<tr>
<th>$m$</th>
<th>Upper bound</th>
<th>Number of variables in the optimization problem</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\omega &lt; 2.3871900$</td>
<td>1</td>
<td>CW (1987)</td>
</tr>
<tr>
<td>2</td>
<td>$\omega &lt; 2.3754770$</td>
<td>3</td>
<td>CW (1987)</td>
</tr>
<tr>
<td>4</td>
<td>$\omega &lt; 2.3729269$</td>
<td>9</td>
<td>Stothers (2010)</td>
</tr>
<tr>
<td>8</td>
<td>$\omega &lt; 2.3729$</td>
<td>29</td>
<td>Vassilevska Williams (2012)</td>
</tr>
<tr>
<td>16</td>
<td>$\omega &lt; 2.3728640$</td>
<td>101</td>
<td>Le Gall (2014)</td>
</tr>
<tr>
<td>32</td>
<td>$\omega &lt; 2.3728639$</td>
<td>373</td>
<td>Le Gall (2014)</td>
</tr>
</tbody>
</table>

Solutions to the optimization problems obtained numerically by *convex optimization*
Conclusion

We constructed a time-efficient implementation of the laser method any tight partitioned tensor for which (lower bounds on) the value of each component is known

Laser-method-based analysis (v2.3) upper bound on $\omega$

convex optimization polynomial time

We applied it to study higher powers of the basic tensor by CW

<table>
<thead>
<tr>
<th>$m$</th>
<th>Upper bound</th>
<th>Number of variables in the optimization problem</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\omega &lt; 2.3871900$</td>
<td>1</td>
<td>CW (1987)</td>
</tr>
<tr>
<td>2</td>
<td>$\omega &lt; 2.3754770$</td>
<td>3</td>
<td>CW (1987)</td>
</tr>
<tr>
<td>4</td>
<td>$\omega &lt; 2.3729269$</td>
<td>9</td>
<td>Stothers (2010)</td>
</tr>
<tr>
<td>8</td>
<td>$\omega &lt; 2.3729$</td>
<td>29</td>
<td>Vassilevska Williams (2012)</td>
</tr>
<tr>
<td>16</td>
<td>$\omega &lt; 2.3728640$</td>
<td>101</td>
<td>Le Gall (2014)</td>
</tr>
<tr>
<td>32</td>
<td>$\omega &lt; 2.3728639$</td>
<td>373</td>
<td>Le Gall (2014)</td>
</tr>
</tbody>
</table>

recent result [Ambainis, Filmus, LG 14]: studying higher powers (using the same approach) cannot give an upper bound better than 2.3725