Codes and Expansions in Algorithms for Matrix Multiplication

Codes & Expansion (CodEx) Seminar
February, 2021

Joshua A. Grochow
Algorithms for Matrix Multiplication
Are
Code-Like Objects
Multiplying matrices

\[
\begin{pmatrix}
  a_{11} & a_{12} & \cdots & a_{1n} \\
  a_{21} & a_{22} & \cdots & a_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{n1} & a_{n2} & \cdots & a_{nn}
\end{pmatrix}
\begin{pmatrix}
  b_{11} \\
  b_{21} \\
  \vdots \\
  b_{n1}
\end{pmatrix}
\begin{pmatrix}
  b_{12} & \cdots & b_{1n} \\
  b_{22} & \cdots & b_{2n} \\
  \vdots & \ddots & \vdots \\
  b_{n2} & \cdots & b_{nn}
\end{pmatrix}
= \begin{pmatrix}
  c_{11} & c_{12} & \cdots & c_{1n} \\
  c_{21} & c_{22} & \cdots & c_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  c_{n1} & c_{n2} & \cdots & c_{nn}
\end{pmatrix}
\]

\(n\) multiplications and \(n - 1\) additions per dot product

\(n^2\) dot products

\[\Rightarrow O(n^3)\] steps
Multiplying matrices

\[
\begin{pmatrix}
  a_{11} & a_{12} & \cdots & a_{1n} \\
  a_{21} & a_{22} & \cdots & a_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{n1} & a_{n2} & \cdots & a_{nn}
\end{pmatrix}
\begin{pmatrix}
  b_{11} & b_{12} & \cdots & b_{1n} \\
  b_{21} & b_{22} & \cdots & b_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  b_{n1} & b_{n2} & \cdots & b_{nn}
\end{pmatrix}
= 
\begin{pmatrix}
  c_{11} & c_{12} & \cdots & c_{1n} \\
  c_{21} & c_{22} & \cdots & c_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  c_{n1} & c_{n2} & \cdots & c_{nn}
\end{pmatrix}
\]

\(n\) multiplications and \(n - 1\) additions per dot product

\(n^2\) dot products

\(\Rightarrow O(n^3)\) steps

Theorem [Klyuyev & Kokovkin-Scherbak ‘65]: Optimal if only allowed to work on rows and columns as a whole.

Theorem [Strassen ‘69]: Can do better! \(O(n^{\log_2 7}) = O(n^{2.81})\).
Strassen’s Algorithm I: A magical 2x2 trick

Ordinary 2x2 product: 8 products, 4 sums

\[ I = (a_{11} + a_{22})(b_{11} + b_{22}) \]
\[ III = (a_{11})(b_{12} - b_{22}) \]
\[ VI = (-a_{11} + a_{21})(b_{11} + b_{12}) \]

Strassen 2x2 product: 7 products, 18 sums

\[
\begin{pmatrix}
1 & 0 & 0 & 1 & -1 & 0 & 1 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & -1 & 1 & 0 & 0 & 1 & 0 \\
\end{pmatrix}
\begin{pmatrix}
I \\
II \\
\vdots \\
VII \\
\end{pmatrix}
\]
Works correctly even if entries are from a noncommutative ring.

Suppose the entries are from $M_n(\mathbb{C})$

\[
\begin{pmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{pmatrix}
\begin{pmatrix}
b_{11} & a_{12} \\
b_{21} & a_{22}
\end{pmatrix}
= 
\begin{pmatrix}
c_{11} & c_{12} \\
c_{21} & c_{22}
\end{pmatrix}
\]
Strassen’s Algorithm II: Recurse

Works correctly even if entries are from a noncommutative ring.

Suppose the entries are from $M_n(\mathbb{C})$

\[
\begin{pmatrix}
  a_{11} & a_{12} \\
  a_{21} & a_{22}
\end{pmatrix}
\begin{pmatrix}
  b_{11} & a_{12} \\
  b_{21} & a_{22}
\end{pmatrix}
= 
\begin{pmatrix}
  c_{11} & c_{12} \\
  c_{21} & c_{22}
\end{pmatrix}
\]

\[
\begin{pmatrix}
  \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\
  \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nn}
\end{pmatrix}
\begin{pmatrix}
  \beta_{11} & \beta_{12} & \cdots & \beta_{1n} \\
  \beta_{21} & \beta_{22} & \cdots & \beta_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  \beta_{n1} & \beta_{n2} & \cdots & \beta_{nn}
\end{pmatrix}
= 
\begin{pmatrix}
  \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} \\
  \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  \gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nn}
\end{pmatrix}
\]

$M_2(M_n(\mathbb{C})) \cong M_{2n}(\mathbb{C})$
Strassen’s Algorithm II: Recurse

Works correctly even if entries are from a noncommutative ring

Suppose the entries are from $M_n(\mathbb{C})$

\[
\begin{pmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{pmatrix} \begin{pmatrix}
b_{11} & a_{12} \\
b_{21} & a_{22}
\end{pmatrix} = \begin{pmatrix}
c_{11} & c_{12} \\
c_{21} & c_{22}
\end{pmatrix}
\]

$M_2(M_n(\mathbb{C})) \cong M_{2n}(\mathbb{C})$

Multiply $2n \times 2n$ matrices using $7$ $n \times n$ products, and $18$ $n \times n$ additions

Let $R(n) = \#$ products ("rank"), $A(n) = \#$ additions to mult two $n \times n$ matrices

$$R(2n) \leq 7R(n)$$

$$A(2n) \leq 18n^2 + 7A(n)$$

$$R(n) \leq n^{\log_2 7}$$

$$A(n) \leq 6(7n^{\log_2 7} - 4n^2) = \Theta(n^{\log_2 7})$$
Theorem [Strassen ‘69]: Can do better! $O(n^{\log_2 7}) = O(n^{2.81})$.

Definition: The exponent of matrix multiplication is
\[ \omega = \inf\{e : MM_n \text{ in } O(n^{e+\varepsilon}) \forall \varepsilon > 0\} \]

Conjecture (folklore): $\omega = 2$.

In principle, $\omega$ depends on the characteristic.
Improvements (in theory)

Current record [Alman-Williams ‘20]: $\omega < 2.372859$
Rephrasing Matrix Multiplication Symmetrically

It is a bilinear map $M_n \otimes M_n \rightarrow M_n$, so we can treat it as an element of $M_n \otimes M_n \otimes M_n^*$, namely

$$T_{MM} = \sum_{ijk} E_{ij} \otimes E_{jk} \otimes E_{ki}$$

$$\langle T_{MM} | A \otimes B \otimes C \rangle = tr(ABC)$$

Note $(AB)_{ik} = tr(ABE_{ki})$
Rephrasing Matrix Multiplication Symmetrically

\[ \sum_{ijk} E_{ij} \otimes E_{jk} \otimes E_{ki} \]

If we write \( M_n \cong U \otimes V^* \), then this lives in

\[ M_n \otimes M_n \otimes M_n^* \]

\[ \cong (U \otimes V^*) \otimes (V \otimes W^*) \otimes (W \otimes U^*) \]

\[ \cong U \otimes (V \otimes V^*) \otimes (W^* \otimes W) \otimes U^* \]

In this decomposition, MM is

\[ \text{Id}_U \otimes \text{Id}_V \otimes \text{Id}_W \]
Rephrasing Matrix Multiplication Symmetrically

\[ \sum_{ijk} E_{ij} \otimes E_{jk} \otimes E_{ki} \]

If we write \( M_n \cong U \otimes V^* \), then this lives in

\[ M_n \otimes M_n \otimes M_n^* \]

\[ \cong (U \otimes V^*) \otimes (V \otimes W^*) \otimes (W \otimes U^*) \]

\[ \cong U \otimes (V^* \otimes V) \otimes (W^* \otimes W) \otimes U^* \]

In this decomposition, MM is

\[ \text{Id}_U \otimes \text{Id}_V \otimes \text{Id}_W \]
Rephrasing Matrix Multiplication
Symmetrically

$$\sum_{ijk} E_{ij} \otimes E_{jk} \otimes E_{ki}$$

If we write $M_n \cong U \otimes V^*$, then this lives in

$$M_n \otimes M_n \otimes M_n^*$$

$$\cong (U \otimes V^*) \otimes (V \otimes W^*) \otimes (W \otimes U^*)$$

$$\cong U \otimes (V^* \otimes V) \otimes (W^* \otimes W) \otimes U^*$$

In this decomposition, MM is

$$Id_U \otimes Id_V \otimes Id_W$$
Rephrasing Matrix Multiplication Symmetrically

\[ \sum_{i,j,k} E_{ij} \otimes E_{jk} \otimes E_{ki} \]

If we write \( M_n \cong U \otimes V^* \), then this lives in

\[ M_n \otimes M_n \otimes M_n^* \]

\[ \cong (U \otimes V^*) \otimes (V \otimes W^*) \otimes (W \otimes U^*) \]

\[ \cong U \otimes (V^* \otimes V) \otimes (W^* \otimes W) \otimes U^* \]

In this decomposition, \( MM \) is

\[ Id_U \otimes Id_V \otimes Id_W \]
Tensor Rank

\[ T_{MM} = \sum_{ijk} E_{ij} \otimes E_{jk} \otimes E_{ki} \]

**Def:** A rank-1 tensor is \( A \otimes B \otimes C \).

A tensor \( t \) has rank \( \leq r \) if it can be written

\[ T = \sum_{i=1}^{r} A_i \otimes B_i \otimes C_i \]

(Generalizes matrix rank)
Def: A tensor $t$ has $\text{rank} \leq r$ if it can be written

$$T = \sum_{i=1}^{r} A_i \otimes B_i \otimes C_i$$

$\text{mult}(T) := \# \text{ of non-scalar multiplications needed to compute } T.$

Thm (classical): $\text{mult}(T) \leq rk(T) \leq 2\text{mult}(T)$
What is an algorithm for computing matrix multiplication?

**Def:** A tensor $t$ has rank $\leq r$ if it can be written

$$T = \sum_{i=1}^{r} A_i \otimes B_i \otimes C_i$$

By an "algorithm for matrix multiplication" we mean

$$\{A_i \otimes B_i \otimes C_i | i = 1, ..., r\}$$

s.t. $T_{MM} = \sum A_i \otimes B_i \otimes C_i$.

This is (somehow) like a code!
Strassen’s Algorithm: A magical 2x2 trick

Strassen 2x2 product: 7 products, 18 sums

\[
\begin{align*}
I &= (a_{11} + a_{22})(b_{11} + b_{22}) \\
III &= (a_{11})(b_{12} - b_{22}) \\
V &= (a_{11} + a_{12})(b_{22}) \\
VII &= (a_{12} - a_{22})(b_{21} + b_{22})
\end{align*}
\]

\[
\begin{pmatrix}
c_{11} \\
c_{12} \\
c_{21} \\
c_{22}
\end{pmatrix} =
\begin{pmatrix}
1 & 0 & 0 & 1 & -1 & 0 & 1 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & -1 & 1 & 0 & 0 & 1 & 0
\end{pmatrix} \begin{pmatrix}
I \\
II \\
\vdots \\
VII
\end{pmatrix}
\]

Strassen: \( rk(T_{MM,2}) \leq 7 \)
Multiplying matrices in practice

\[
\begin{pmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{nn}
\end{pmatrix}
\begin{pmatrix}
b_{11} & b_{12} & \cdots & b_{1n} \\
b_{21} & b_{22} & \cdots & b_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
b_{n1} & b_{n2} & \cdots & b_{nn}
\end{pmatrix}
= 
\begin{pmatrix}
c_{11} & c_{12} & \cdots & c_{1n} \\
c_{21} & c_{22} & \cdots & c_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
c_{n1} & c_{n2} & \cdots & c_{nn}
\end{pmatrix}
\]

Used in scientific computing*, graphics, GPUs, combinatorial algorithms, other algebraic algorithms, deep learning

Matrix multiplication can even be the bottleneck
In practice

- Often have sparse or structured matrices → better algorithms available
- Even with dense, unstructured matrices, memory and communication are more frequently the bottleneck than number of operations
Multiplying matrices in practice

In practice

• Often have sparse or structured matrices → better algorithms available

• Even with dense, unstructured matrices, memory and communication are more frequently the bottleneck than number of operations

• Strassen’s algorithm actually used in practice

• Other algorithms today aren’t, but future ones could be!
Multiplying matrices

\[
\begin{pmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{nn}
\end{pmatrix}
\begin{pmatrix}
b_{11} & b_{12} & \cdots & b_{1n} \\
b_{21} & b_{22} & \cdots & b_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
b_{n1} & b_{n2} & \cdots & b_{nn}
\end{pmatrix} =
\begin{pmatrix}
c_{11} & c_{12} & \cdots & c_{1n} \\
c_{21} & c_{22} & \cdots & c_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
c_{n1} & c_{n2} & \cdots & c_{nn}
\end{pmatrix}
\]

Used in scientific computing*, graphics, GPUs, combinatorial algorithms, other algebraic algorithms, deep learning
Matrix multiplication can even be the bottleneck

More importantly: gives us insight into the nature of computing & complexity!
How did Strassen find this?

Trying to prove it couldn’t be done mod 2, by hand, (clever) brute force.

Found solution mod 2.

Figured out signs to work over $\mathbb{Z}$. 
Mystery:
Math behind the magic?

Where does the magical $2 \times 2$ trick “really” come from?
Math behind the magic I
Clausen’s Construction

Multiplication table

<table>
<thead>
<tr>
<th></th>
<th>$E_{11}$</th>
<th>$E_{12}$</th>
<th>$E_{21}$</th>
<th>$E_{22}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_{11}$</td>
<td>$E_{11}$</td>
<td>$E_{12}$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$E_{21}$</td>
<td>$E_{21}$</td>
<td>$E_{22}$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$E_{12}$</td>
<td>0</td>
<td>0</td>
<td>$E_{11}$</td>
<td>$E_{12}$</td>
</tr>
<tr>
<td>$E_{22}$</td>
<td>0</td>
<td>0</td>
<td>$E_{21}$</td>
<td>$E_{22}$</td>
</tr>
</tbody>
</table>

Leads to tensor rank decomposition

$$T_{MM,2} = \sum E_{11} \otimes E_{11} \otimes E_{11} + E_{21} \otimes E_{11} \otimes E_{21} + \cdots$$

$$+ E_{22} \otimes E_{22} \otimes E_{22}$$
Math behind the magic I
Clausen’s Construction

Multiplication table

<table>
<thead>
<tr>
<th></th>
<th>$E_{11}$</th>
<th>$E_{12}$</th>
<th>$E_{21}$</th>
<th>$E_{22}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_{11}$</td>
<td>$E_{11}$</td>
<td>$E_{12}$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$E_{21}$</td>
<td>$E_{21}$</td>
<td>$E_{22}$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$E_{12}$</td>
<td>0</td>
<td>0</td>
<td>$E_{11}$</td>
<td>$E_{12}$</td>
</tr>
<tr>
<td>$E_{22}$</td>
<td>0</td>
<td>0</td>
<td>$E_{21}$</td>
<td>$E_{22}$</td>
</tr>
</tbody>
</table>

Leads to tensor rank decomposition

$$T_{MM,2} = E_{11} \otimes E_{11} \otimes E_{11} + E_{21} \otimes E_{11} \otimes E_{21} + \cdots + E_{22} \otimes E_{22} \otimes E_{22}$$
Math behind the magic I
Clausen’s Construction

Multiplication table

\[
\begin{array}{cccc}
E_{11} & E_{12} & E_{21} & E_{22} \\
E_{11} & E_{11} & E_{12} & 0 & 0 \\
E_{21} & E_{21} & E_{22} & 0 & 0 \\
E_{12} & 0 & 0 & E_{11} & E_{12} \\
E_{22} & 0 & 0 & E_{21} & E_{22} \\
\end{array}
\]

Leads to tensor rank decomposition

\[
T_{MM,2} = E_{11} \otimes E_{11} \otimes E_{11} + E_{21} \otimes E_{11} \otimes E_{21} + \cdots + E_{22} \otimes E_{22} \otimes E_{22}
\]
Math behind the magic I
Clausen’s Construction

Multiplication table

<table>
<thead>
<tr>
<th></th>
<th>$E_{11}$</th>
<th>$E_{12}$</th>
<th>$E_{21}$</th>
<th>$E_{22}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_{11}$</td>
<td>$E_{11}$</td>
<td>$E_{12}$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$E_{21}$</td>
<td>$E_{21}$</td>
<td>$E_{22}$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$E_{12}$</td>
<td>0</td>
<td>0</td>
<td>$E_{11}$</td>
<td>$E_{12}$</td>
</tr>
<tr>
<td>$E_{22}$</td>
<td>0</td>
<td>0</td>
<td>$E_{21}$</td>
<td>$E_{22}$</td>
</tr>
</tbody>
</table>

Leads to tensor rank decomposition

$$T_{MM,2} =$$

$$E_{11} \otimes E_{11} \otimes E_{11} + E_{21} \otimes E_{11} \otimes E_{21} + \cdots$$

$$+ E_{22} \otimes E_{22} \otimes E_{22}$$
Consider the action of $S_3$ on $\mathbb{C}^2$ given by

$$(12) \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad (123) \mapsto \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}$$

Let $X_0 = Y_0 = Id_2$

Let $X_1 = E_{11}$.

Let $X_1, X_2, X_3$ be the orbit of $X_1$ under conjugation by $C_3 \leq S_3$.

Let $Y_i = (12)X_i(12)$. 

Clausen ‘88 thesis

See Bürgisser-Clausen-Shokrollahi Ch. 1
Claim: \( \{X_0, X_1, X_2, X_3\} \) is a basis of \( M_2 \). (So is \( \{Y_0, Y_1, Y_2, Y_3\} \).) Their multiplication table is

\[
\begin{array}{cccc}
Y_0 & Y_1 & Y_2 & Y_3 \\
X_0 & Y_0 & Y_1 & Y_2 & Y_3 \\
X_1 & X_1 & 0 & Y_2 & X_1 \\
X_2 & X_2 & Y_1 & X_2 & 0 \\
X_3 & X_3 & X_3 & 0 & Y_3
\end{array}
\]

Clausen ‘88 thesis
See Bürgisser-Clausen-Shokrollahi Ch. 1
Claim: \( \{X_0, X_1, X_2, X_3\} \) is a basis of \( M_2 \). (So is \( \{Y_0, Y_1, Y_2, Y_3\} \).)

Their multiplication table is

\[
\begin{array}{cccc}
Y_0 & Y_1 & Y_2 & Y_3 \\
X_0 & Y_0 & Y_1 & Y_2 & Y_3 \\
X_1 & X_1 & 0 & Y_2 & X_1 \\
X_2 & X_2 & Y_1 & X_2 & 0 \\
X_3 & X_3 & X_3 & 0 & Y_3
\end{array}
\]

The rank decomposition has terms that combine:

\[
T_{MM,2} = X_0 \otimes Y_0 \otimes Y_0 + X_1 \otimes (Y_0 + Y_3) \otimes X_1 + \ldots + X_3 \otimes (Y_0 + Y_1) \otimes X_3
\]

Clausen ‘88 thesis

See Bürgisser-Clausen-Shokrollahi Ch. 1
Claim: \( \{X_0, X_1, X_2, X_3\} \) is a basis of \( M_2 \). (So is \( \{Y_0, Y_1, Y_2, Y_3\} \). Their multiplication table is

\[
\begin{array}{cccc}
Y_0 & Y_1 & Y_2 & Y_3 \\
X_0 & Y_0 & Y_1 & Y_2 & Y_3 \\
X_1 & X_1 & 0 & Y_2 & X_1 \\
X_2 & X_2 & Y_1 & X_2 & 0 \\
X_3 & X_3 & X_3 & 0 & Y_3
\end{array}
\]

The rank decomposition has terms that combine:

\[
T_{MM,2} = X_0 \otimes Y_0 \otimes Y_0 + X_1 \otimes (Y_0 + Y_3) \otimes X_1 + \cdots + X_3 \otimes (Y_0 + Y_1) \otimes X_3
\]
Claim: \( \{X_0, X_1, X_2, X_3\} \) is a basis of \( M_2 \). (So is \( \{Y_0, Y_1, Y_2, Y_3\} \).

Their multiplication table is

\[
\begin{array}{cccc}
Y_0 & Y_1 & Y_2 & Y_3 \\
X_0 & Y_0 & Y_1 & Y_2 & Y_3 \\
X_1 & X_1 & 0 & Y_2 & X_1 \\
X_2 & X_2 & Y_1 & X_2 & 0 \\
X_3 & X_3 & X_3 & 0 & Y_3 \\
\end{array}
\]

The rank decomposition has terms that combine:

\[
T_{MM,2} =
X_0 \otimes Y_0 \otimes Y_0 + X_1 \otimes (Y_0 + Y_3) \otimes X_1 + \cdots \\
+ X_3 \otimes (Y_0 + Y_1) \otimes X_3
\]

Clausen ‘88 thesis

See Bürgisser-Clausen-Shokrollahi Ch. 1
Math behind the magic I
Clausen’s Construction

Claim: \{X_0, X_1, X_2, X_3\} is a basis of \(M_2\). (So is \{Y_0, Y_1, Y_2, Y_3\}. Their multiplication table is

\[
\begin{array}{cccc}
Y_0 & Y_1 & Y_2 & Y_3 \\
X_0 & Y_0 & Y_1 & Y_2 & Y_3 \\
X_1 & X_1 & 0 & Y_2 & X_1 \\
X_2 & X_2 & Y_1 & X_2 & 0 \\
X_3 & X_3 & X_3 & 0 & Y_3
\end{array}
\]

The rank decomposition has terms that combine:

\[
T_{MM,2} =
X_0 \otimes Y_0 \otimes Y_0 + X_1 \otimes (Y_0 + Y_3) \otimes X_1 + \cdots \\
+ X_3 \otimes (Y_0 + Y_1) \otimes X_3
\]

Clausen ‘88 thesis
See Bürgisser-Clausen-Shokrollahi Ch. 1
Claim: $\{X_0,X_1,X_2,X_3\}$ is a basis of $M_2$. (So is $\{Y_0,Y_1,Y_2,Y_3\}$.

Their multiplication table is

\[
\begin{array}{cccc}
Y_0 & Y_1 & Y_2 & Y_3 \\
X_0 & Y_0 & Y_1 & Y_2 & Y_3 \\
X_1 & X_1 & 0 & Y_2 & X_1 \\
X_2 & X_2 & Y_1 & X_2 & 0 \\
X_3 & X_3 & X_3 & 0 & Y_3 \\
\end{array}
\]

The rank decomposition has terms that combine:

\[
T_{MM,2} = X_0 \otimes Y_0 \otimes Y_0 + X_1 \otimes (Y_0 + Y_3) \otimes X_1 + \cdots + X_3 \otimes (Y_0 + Y_1) \otimes X_3
\]

Only 7 terms! 

Clausen ‘88 thesis See Bürgisser-Clausen-Shokrollahi Ch. 1
Consider the action of $S_3$ on $\mathbb{C}^2$ given by

$$(12) \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad (123) \mapsto \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}$$

Let $X_0 = Y_0 = \text{Id}_2$

Let $X_1 = E_{11}$.

Let $X_1, X_2, X_3$ be the orbit of $X_1$ under conjugation by $C_3 \leq S_3$.

Let $Y_i = (12)X_i(12)$. 

Reminiscent of construction of Golay code

Clausen ‘88 thesis

See Bürgisser-Clausen-Shokrollahi Ch. 1
Similar (if you squint)
Construction of Golay code
Similar (if you squint)
Construction of Golay code

Thompson, From Error-Correcting Codes Through Sphere Packings to Simple Groups, MAA 1984
Similar (if you squint)
Construction of Golay code

Thompson, *From Error-Correcting Codes Through Sphere Packings to Simple Groups*, MAA 1984
Aside

Second Connection With Codes
Second connection with codes
Lempel-Winograd ’76 (not squinty)

Strassen 2x2 product: 7 products, 18 sums

\[ I = (a_{11} + a_{22})(b_{11} + b_{22}) \]
\[ III = (a_{11})(b_{12} - b_{22}) \]
\[ VI = (-a_{11} + a_{21})(b_{11} + b_{12}) \]
\[ VII = (a_{12} - a_{22})(b_{21} + b_{22}) \]
\[ II = (a_{21} + a_{22})(b_{11}) \]
\[ IV = (a_{22})(-b_{11} + b_{21}) \]
\[ V = (a_{11} + a_{12})(b_{22}) \]

\[
\begin{pmatrix}
 c_{11} \\
 c_{12} \\
 c_{21} \\
 c_{22}
\end{pmatrix} = \begin{pmatrix}
 1 & 0 & 0 & 1 & -1 & 0 & 1 \\
 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
 1 & -1 & 1 & 0 & 0 & 1 & 0 \\
\end{pmatrix} \begin{pmatrix}
 I \\
 II \\
 \vdots \\
 VII
\end{pmatrix} \]
Second connection with codes
Lempel-Winograd ’76 (not squinty)

\[
\begin{pmatrix}
  c_{11} \\
  c_{12} \\
  c_{21} \\
  c_{22}
\end{pmatrix}
= \begin{pmatrix}
  1 & 0 & 0 & 1 & -1 & 0 & 1 \\
  0 & 0 & 1 & 0 & 1 & 0 & 0 \\
  0 & 1 & 0 & 1 & 0 & 0 & 0 \\
  1 & -1 & 1 & 0 & 0 & 1 & 0
\end{pmatrix} \begin{pmatrix}
  I \\
  II \\
  \vdots \\
  VII
\end{pmatrix}
\]

Thm [Lempel-Winograd]: Rank \( n \) decompositions of a tensor with \( k \) outputs correspond to \([n, k, d]\) codes with \( d \geq \) (something having to do with the tensor).

For \( n \times n \) matrix multiplication, these are \([n^\omega, n^2, n]\) codes.
Back to the Story
Matrix multiplication is characterized by its symmetries (important in Geometric Complexity Theory approach to P versus NP).

Let $\mu: M_n \otimes M_n \to M_n$ be matrix multiplication. Then:

$$\mu(A, B) = X^{-1} \mu(XAY^{-1}, YBZ^{-1})Z$$

for all $X, Y, Z \in GL_n$.

Any bilinear map $f: M_n \otimes M_n \to M_n$ such that

$$f(A, B) = X^{-1} f(XAY^{-1}, YBZ^{-1})Z$$

for all $X, Y, Z \in GL_n$ is a scalar multiple of $\mu$.

Our starting point: For $V$ an irrep of $G$, write $M_n = V \otimes V^*$, then invariance under $G^3$ (rather than $GL^3_n$) suffices!
\[ \sum_{ijk} E_{ij} \otimes E_{jk} \otimes E_{ki} \]

If we write \( M_n \cong U \otimes V^* \), then this lives in
\[
M_n \otimes M_n \otimes M_n^* \cong (U \otimes V^*) \otimes (V \otimes W^*) \otimes (W \otimes U^*) \cong U \otimes (V \otimes V^*) \otimes (W^* \otimes W) \otimes U^*
\]

In this decomposition, \( MM \) is
\[
\text{Id}_U \otimes \text{Id}_V \otimes \text{Id}_W
\]
2×2 matrix multiplication
Using group orbits

G: Finite group, acting irreducibly on $\mathbb{C}^n$.
S: G-orbit of in $\mathbb{C}^n$. Then (Schur’s Lemma)

$$\frac{n}{|S|} \sum_{v \in S} |v\rangle \langle v| = Id$$

Then we have

$$\frac{n^3}{|S|^3} \sum_{u,v,w \in S} |u\rangle \langle u| \otimes |v\rangle \langle v| \otimes |w\rangle \langle w| = Id \otimes Id \otimes Id$$

$$\frac{n^3}{|S|^3} \sum_{u,v,w \in S} |u\rangle \langle v| \otimes |v\rangle \langle w| \otimes |w\rangle \langle u| = MM$$

Suppose $S \subseteq \mathbb{C}^n$ satisfies
\[
\frac{1}{|S|} \sum_{v \in S} |v\rangle\langle v| = \frac{1}{n} Id
\]

Then we have
\[
\frac{n^3}{|S|^3} \sum_{u,v,w \in S} |u\rangle\langle u| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w| = Id \otimes Id \otimes Id
\]
\[
\frac{n^3}{|S|^3} \sum_{u,v,w \in S} |u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle u| = MM
\]

Proof

Suppose $S \subseteq \mathbb{C}^n$ satisfies

$$\frac{1}{|S|} \sum_{v \in S} |v\rangle\langle v| = \frac{1}{n} Id$$

Then we have

$$\frac{n^3}{|S|^3} \sum_{u,v,w \in S} |u\rangle\langle u| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w| = Id \otimes Id \otimes Id$$

$$\frac{n^3}{|S|^3} \sum_{u,v,w \in S} |u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle u| = MM$$

|S|^3 summands

G. & Moore arXiv:1612.01527
and arXiv:1708.09398
Proof

Consider

\[ \sum_{u,v,w \in S} |u\rangle\langle v - u| \otimes |v\rangle\langle w - v| \otimes |w\rangle\langle u - w| \]

Proof

Consider
\[
\sum_{u,v,w \in S} |u\rangle\langle v - u| \otimes |v\rangle\langle w - v| \otimes |w\rangle\langle u - w|
\]

Why consider this?
• Looks kinda like matrix multiplication. Almost \(G^3\)-invariant.
• Only need to sum of triples of distinct \(u, v, w \in S\).
  \(\rightarrow |S|(|S| - 1)(|S| - 2)\) summands < \(|S|^3\)

Proof

\[ \sum_{u,v,w \in S} |u\rangle\langle v - u| \otimes |v\rangle\langle w - v| \otimes |w\rangle\langle u - w| \]

Expand out, giving 4 kinds of terms:

1. Matched \(-|u\rangle\langle u| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|\)
2. 1 mismatch \(|u\rangle\langle v| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|\)
3. 2 mismatches \(|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle w|\)
4. 3 mismatches \(|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle u|\)

Proof

\[
\sum_{u,v,w \in S} |u\rangle\langle v - u| \otimes |v\rangle\langle w - v| \otimes |w\rangle\langle u - w|
\]

Expand out, giving 4 kinds of terms:

1. Matched \(-|u\rangle\langle u| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|\)
2. 1 mismatch \(|u\rangle\langle v| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|\)
3. 2 mismatches \(|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle w|\)
4. 3 mismatches \(|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle u|\)

Proof

\[ \sum_{u,v,w \in S} |u\rangle\langle v - u| \otimes |v\rangle\langle w - v| \otimes |w\rangle\langle u - w| \]

Expand out, giving 4 kinds of terms:

1. Matched \(-|u\rangle\langle u| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|\)
2. 1 mismatch \(|u\rangle\langle v| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|\)
3. 2 mismatches \(|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle w|\)
4. 3 mismatches \(|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle u|\)

\[ \sum_{u,v,w \in S} |u \rangle \langle v - u| \otimes |v \rangle \langle w - v| \otimes |w \rangle \langle u - w| \]

Expand out, giving 4 kinds of terms:

1. Matched \(-|u \rangle \langle u| \otimes |v \rangle \langle v| \otimes |w \rangle \langle w|\)
2. 1 mismatch \(|u \rangle \langle v| \otimes |v \rangle \langle v| \otimes |w \rangle \langle w|\)
3. 2 mismatches \(|u \rangle \langle v| \otimes |v \rangle \langle w| \otimes |w \rangle \langle w|\)
4. 3 mismatches \(|u \rangle \langle v| \otimes |v \rangle \langle w| \otimes |w \rangle \langle u|\)

Proof

\[
\sum_{u,v,w \in S} |u\rangle\langle v - u| \otimes |v\rangle\langle w - v| \otimes |w\rangle\langle u - w|
\]

Expand out, giving 4 kinds of terms:

1. Matched \(-|u\rangle\langle u| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w| \rightarrow Id \otimes^3\)

2. 1 mismatch \(|u\rangle\langle v| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|\)

3. 2 mismatches \(|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle w|\)

4. 3 mismatches \(|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle u|\)

G. & Moore arXiv:1612.01527
and arXiv:1708.09398
Proof

\[\sum_{u,v,w \in S} |u\rangle\langle v - u| \otimes |v\rangle\langle w - v| \otimes |w\rangle\langle u - w|\]

Expand out, giving 4 kinds of terms:

1. Matched \(-|u\rangle\langle u| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w| \rightarrow Id \otimes^3\)

2. 1 mismatch \(|u\rangle\langle v| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|\)

3. 2 mismatches \(|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle w|\)

4. 3 mismatches \(|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle u| \rightarrow MM\)

G. & Moore \hspace{1em} arXiv:1612.01527
and \hspace{1em} arXiv:1708.09398
Proof

\[\sum_{u,v,w \in S} |u\rangle\langle v - u| \otimes |v\rangle\langle w - v| \otimes |w\rangle\langle u - w|\]

Expand out, giving 4 kinds of terms:

1. Matched \(-|u\rangle\langle u| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|\) \(\rightarrow Id^{\otimes 3}\)
2. 1 mismatch \(|u\rangle\langle v| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|\) \(\rightarrow 0?\)
3. 2 mismatches \(|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle w|\) \(\rightarrow 0?\)
4. 3 mismatches \(|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle u|\) \(\rightarrow MM\)

G. & Moore arXiv:1612.01527
and arXiv:1708.09398
Proof

\[ \sum_{u,v,w \in S} |u\rangle\langle v - u| \otimes |v\rangle\langle w - v| \otimes |w\rangle\langle u - w| \]

Expand out, giving 4 kinds of terms:

1. Matched \(-|u\rangle\langle u| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w| \rightarrow Id \otimes^3\)
2. 1 mismatch \(|u\rangle\langle v| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w| \rightarrow 0?\)
3. 2 mismatches \(|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle w| \rightarrow 0?\)
4. 3 mismatches \(|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle u| \rightarrow MM\)

Proof

If

\[ \sum_{u \in S} u = 0 \quad (*) \]

Then those terms vanish, as desired.

Given \( u \in \mathbb{C}^n \), (*) is its projection onto trivial rep.

\( \rightarrow \) If \( \mathbb{C}^n \) is a nontrivial irrep, the sum must be 0

Proof

If $S$ is a $G$-orbit in a nontrivial irrep, then

$$MM = Id^3 + \frac{n^3}{|S|^3} \sum_{u,v,w \in S} |u\rangle\langle v - u| \otimes |v\rangle\langle w - v| \otimes |w\rangle\langle u - w|$$
Proof

If

\[ \sum_{u \in S} u = 0 \]

\[ \frac{1}{|S|} \sum_{v \in S} |v\rangle\langle v| = \frac{1}{n} \text{id} \]

Then

\[ Id^{\otimes 3} + \frac{n^3}{|S|^3} \sum_{u,v,w \in S} |u\rangle\langle v-u| \otimes |v\rangle\langle w-v| \otimes |w\rangle\langle u-w| \]

\[ = MM \]

Unitary 2-Design!

Theorem [G.-Moore]: For $S \subseteq \mathbb{C}^n$ a unitary 2-design, $s = |S|$, $n \times n$ matrices can be multiplied using at most $s(s - 1)(s - 2) + 1$ multiplications.

The math behind the magic II

**Theorem [G.-Moore]:** For $S \subseteq \mathbb{C}^n$ a unitary 2-design, $s = |S|$, $n \times n$ matrices can be multiplied using at most $s(s - 1)(s - 2) + 1$ multiplications.

**Observe:** If $G$ finite group, $V$ nontrivial irrep, $\nu \in V$ w/ $|\nu| = 1$, then the orbit $G\nu$ is a unitary 2-design.

Theorem [G.-Moore]: For $S \subseteq \mathbb{C}^n$ a unitary 2-design, $s = |S|$, $n \times n$ matrices can be multiplied using at most $s(s - 1)(s - 2) + 1$ multiplications.

The action of $S_3$ on $\mathbb{C}^2$ has an orbit of size 3 (equilateral triangle).

Corollary [G.-Moore]: $3(3 - 1)(3 - 2) + 1 = 7$. 

The math behind the magic II

Theorem [G.-Moore]: For \( S \subseteq \mathbb{C}^n \) a unitary 2-design, \( s = |S| \), \( n \times n \) matrices can be multiplied using at most \( s(s - 1)(s - 2) + 1 \) multiplications.

The action of \( S_{n+1} \) on \( \mathbb{C}^n \) has an orbit of size \( n+1 \). Smallest possible, since must sum to zero.

Corollary [G.-Moore]: \( (n + 1)n(n - 1) + 1 = n^3 - n + 1 < n^3 \).

Okay, but is that the same as Strassen’s algorithm?

- Yes, by a theorem of de Groote ‘78
- But…we don’t care! It gives a conceptual explanation of the upper bound of 7.
Okay, but is that the same as Strassen’s algorithm?

- Yes, by a theorem of de Groote ‘78
- But…we don’t care! It gives a conceptual explanation of the upper bound of 7.

Third Connection: Uniqueness, similar to uniqueness of Golay code
Summary of
The math behind the magic

Gave conceptual explanation of Strassen’s 7.

**Open:** Find a similar explanation that works over arbitrary rings (as Strassen’s algorithm does; ours needs $\frac{1}{2}$, $\frac{1}{3}$, and $\sqrt{3}$). (Arithmetic geometry)

Further Ideas:

- Use of unitary $d$-designs for $d > 2$? (cf. iterated matrix multiplication)
- Other symmetric algorithms? See also [Burichenko](#); [Chiantini-Ikenmeyer-Landsberg-Ottaviani](#).

Other explanations
Of Strassen’s 2×2 algorithm

Gastinel 1971
Yuval 1978
Chatelin 1986
Clausen 1988 (thesis)
Alekseyev 1996
Gates & Kreinovich 2000 (Bull. EATCS)
Paterson 2009 (Valiant’s 60th)
Minz 2015
Chiantini, Ikenmeyer, Landsberg, Ottaviani 2016
Ikenmeyer & Lysikov 2017

Uses/reveals symmetries
Additional connections
With code-like things

• Some algorithms highly symmetric, some not. Symmetry, optimality, and uniqueness.
• Arithmetic progression-free sets (Coppersmith-Winograd ‘90)
• Cohn-Umans ‘03 group-theoretic approach has a packing problem at its core
• Coherent configurations (Cohn-Umans ‘13)
• Can phrase MM as a minimization problem. Similar (squint) to spherical codes / the Thomson problem
• Combine algorithms for small MM into algorithms for larger MM (e.g. Hopcroft-Kerr ‘71, Strassen’s laser method). Similar (squint!) to construction of Leech lattice
Additional references

(Note hyperlinks throughout the document – the ones here are additional)

Symmetries in algorithms for matrix multiplication
• Burichenko arXiv:1408.6273
• Landsberg & Ryder, Exp. Math. 2017
• Landsberg & Michałek, arXiv:1601.08229

Cap sets & Cohn-Umans approach
• Blasiak, Church, Cohn, G. Naslund, Sawin, Umans, Disc. Analysis 2017
• Blasiak, Church, Cohn, G., Umans arXiv:1712.02302