Fast Matrix Multiplication

Michael Schmid

31.05.2021

Big $\mathcal{O}$ notation

- Time complexity of an algorithm

Big $\mathcal{O}$ notation

- Time complexity of an algorithm
- How many multiplications in a function

Big $\mathcal{O}$ notation

- Time complexity of an algorithm
- How many multiplications in a function
- Drop Constants

Big $\mathcal{O}$ notation

Algorithm 1 Foo 1

```
function FOO(a, b)
    return a + b
```

Big $\mathcal{O}$ notation

Algorithm 2 Foo 1

```
function FOO(a, b)
    return a + b
```

$\mathcal{O}(1)$

Big $\mathcal{O}$ notation

Algorithm 3 Foo 2

```
function FOO(a, b)
    x ← a + b
    y ← a · b
    return x + y
```

Big $\mathcal{O}$ notation

Algorithm 4 Foo 2

```
function FOO(a, b)
    x ← a + b
    y ← a · b
    return x + y
```

$$\mathcal{O}(1) + \mathcal{O}(1) = 2\mathcal{O}(1) = \mathcal{O}(1)$$

Big $\mathcal{O}$ notation

Algorithm 5 Foo 3

function FOO($\mathbf{A}, \mathbf{B}$,n)
    $sum \leftarrow 0$
    for $i = 0, 1, 2 \ldots, n$ **do**
        $sum \leftarrow sum + A[i] \cdot B[i]$
    return sum

Big $\mathcal{O}$ notation

Algorithm 6 Foo 3

function FOO(**A**, **B**,n)
    $sum \leftarrow 0$
    for $i = 0, 1, 2 \ldots, n$ **do**
        $sum \leftarrow sum + A[i] \cdot B[i]$
    return sum

$\mathcal{O}(n)$

Big $\mathcal{O}$ notation

Algorithm 7 Foo 1

function $\textsc{Foo}(\mathbf{A}, \mathbf{B}, \mathrm{n})$
  $sum \leftarrow 0$
  for $i = 0, 1, 2 \ldots, n$ **do**
    for $j = 0, 1, 2 \ldots, n$ **do**
      $sum \leftarrow sum + A[i] \cdot B[j]$
  return sum

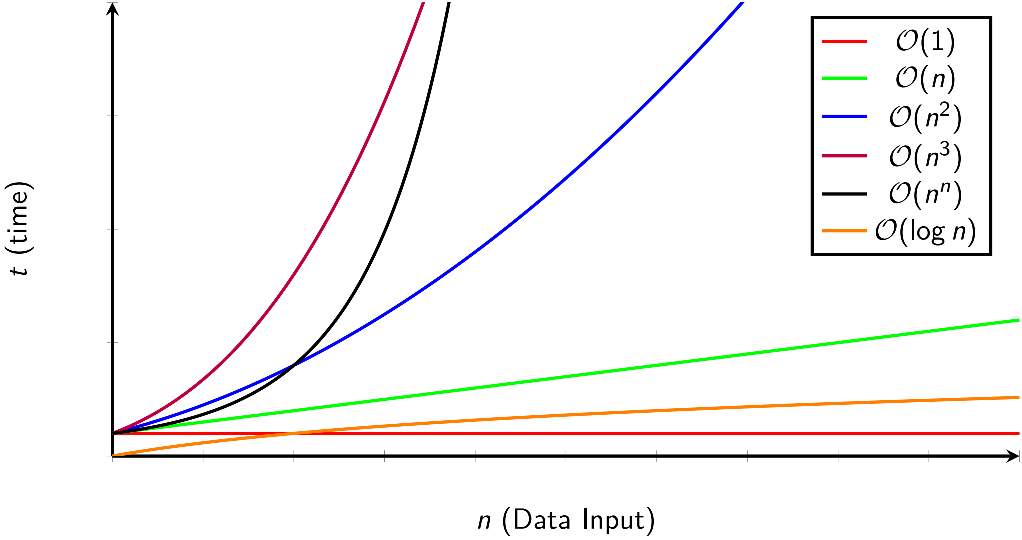
Big $\mathcal{O}$ notation

Algorithm 8 Foo 1

```
function FOO(A, B,n)
    sum ← 0
    for i = 0, 1, 2 . . . , n do
        for j = 0, 1, 2 . . . , n do
            sum ← sum + A[i] · B[j]
    return sum
```

$\mathcal{O}(n^2)$

Big $\mathcal{O}$ notation



Strassen's Algorithm

$$\mathbf{AB} = \mathbf{C}$$

Strassen's Algorithm

$$\mathbf{AB} = \mathbf{C}$$

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

Strassen's Algorithm

$$\mathbf{AB} = \mathbf{C}$$

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

$$C_{11} = A_{11} \cdot B_{11} + A_{12} \cdot B_{21}$$
$$C_{12} = A_{11} \cdot B_{12} + A_{12} \cdot B_{22}$$
$$C_{21} = A_{21} \cdot B_{11} + A_{22} \cdot B_{21}$$
$$C_{22} = A_{21} \cdot B_{12} + A_{22} \cdot B_{22}$$

Algorithm

Algorithm 9 Square Matrix Multiplication

```
function MM(A, B, C)
    sum ← 0
    n ← columns(A) == rows(B)
    m ← rows(A)
    p ← columns(B)
    for i = 0, 1, 2 ..., m − 1 do
        for j = 0, 1, 2 ..., p − 1 do
            sum ← 0
            for k = 0, 1, 2 ..., n − 1 do
                sum ← sum + A[i][k] · B[k][j]
            C[i][j] ← sum
    return C
```

$$\begin{bmatrix} B_{1,1} & \cdots & B_{1,j} & \cdots & B_{1,p} \\ \vdots & & \vdots & & \vdots \\ B_{k,1} & \cdots & B_{k,j} & \cdots & B_{k,p} \\ \vdots & & \vdots & & \vdots \\ B_{n,1} & \cdots & B_{n,j} & \cdots & B_{n,p} \end{bmatrix}$$

$$\begin{bmatrix} A_{1,1} & \cdots & A_{1,k} & \cdots & A_{1,n} \\ \vdots & & \vdots & & \vdots \\ A_{i,1} & \cdots & A_{i,k} & \cdots & A_{i,n} \\ \vdots & & \vdots & & \vdots \\ A_{m,1} & \cdots & A_{m,k} & \cdots & A_{m,n} \end{bmatrix} \quad \begin{bmatrix} C_{1,1} & \cdots & C_{1,j} & \cdots & C_{1,p} \\ \vdots & & \vdots & & \vdots \\ C_{i,1} & \cdots & C_{i,j} & \cdots & C_{i,p} \\ \vdots & & \vdots & & \vdots \\ C_{m,1} & \cdots & C_{m,j} & \cdots & C_{m,p} \end{bmatrix}$$

Strassen's Algorithm

$$
\begin{aligned}
\mathsf{I} &= (A_{11} + A_{22}) \cdot (B_{11} + B_{22}) \\
\mathsf{II} &= (A_{21} + A_{22}) \cdot B_{11} \\
\mathsf{III} &= A_{11} \cdot (B_{12} - B_{22}) \\
\mathsf{IV} &= A_{22} \cdot (-B_{11} + B_{21}) \\
\mathsf{V} &= (A_{11} + A_{12}) \cdot B_{22} \\
\mathsf{VI} &= (-A_{11} + A_{21}) \cdot (B_{11} + B_{12}) \\
\mathsf{VII} &= (A_{12} - A_{22}) \cdot (B_{21} + B_{22})
\end{aligned}
$$

Strassen's Algorithm

$$\begin{aligned}
\mathrm{I} &= (A_{11} + A_{22}) \cdot (B_{11} + B_{22}) \\
\mathrm{II} &= (A_{21} + A_{22}) \cdot B_{11} \\
\mathrm{III} &= A_{11} \cdot (B_{12} - B_{22}) \\
\mathrm{IV} &= A_{22} \cdot (-B_{11} + B_{21}) \\
\mathrm{V} &= (A_{11} + A_{12}) \cdot B_{22} \\
\mathrm{VI} &= (-A_{11} + A_{21}) \cdot (B_{11} + B_{12}) \\
\mathrm{VII} &= (A_{12} - A_{22}) \cdot (B_{21} + B_{22})
\end{aligned}$$

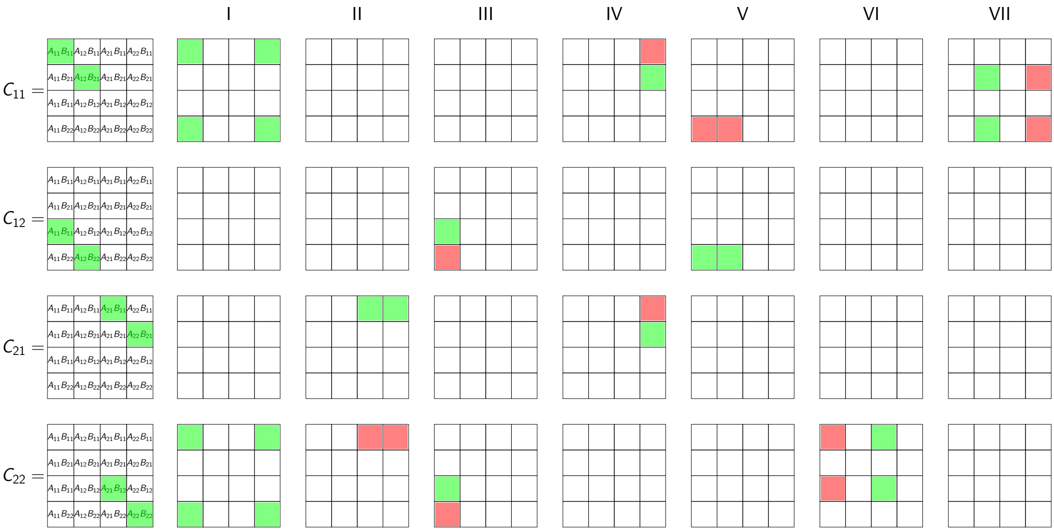
$$\begin{aligned}
C_{11} &= \mathrm{I} + \mathrm{IV} - \mathrm{V} + \mathrm{VII} \\
C_{21} &= \mathrm{II} + \mathrm{IV} \\
C_{12} &= \mathrm{III} + \mathrm{V} \\
C_{22} &= \mathrm{I} + \mathrm{III} - \mathrm{II} + \mathrm{VI}
\end{aligned}$$

Strassen's Algorithm

$$
\begin{aligned}
\mathrm{I} &= (A_{11} + A_{22}) \cdot (B_{11} + B_{22}) \\
\mathrm{II} &= (A_{21} + A_{22}) \cdot B_{11} \\
\mathrm{III} &= A_{11} \cdot (B_{12} - B_{22}) \\
\mathrm{IV} &= A_{22} \cdot (-B_{11} + B_{21}) \\
\mathrm{V} &= (A_{11} + A_{12}) \cdot B_{22} \\
\mathrm{VI} &= (-A_{11} + A_{21}) \cdot (B_{11} + B_{12}) \\
\mathrm{VII} &= (A_{12} - A_{22}) \cdot (B_{21} + B_{22})
\end{aligned}
$$

$$
\begin{aligned}
C_{11} &= \mathrm{I} + \mathrm{IV} - \mathrm{V} + \mathrm{VII} \\
C_{21} &= \mathrm{II} + \mathrm{IV} \\
C_{12} &= \mathrm{III} + \mathrm{V} \\
C_{22} &= \mathrm{I} + \mathrm{III} - \mathrm{II} + \mathrm{VI}
\end{aligned}
$$

$$
\begin{aligned}
C_{11} &= (A_{11} + A_{22}) \cdot (B_{11} + B_{22}) + A_{22} \cdot (-B_{11} + B_{21}) - (A_{11} + A_{12}) \cdot B_{22} + (A_{12} - A_{22}) \cdot (B_{21} + B_{22}) \\
C_{11} &= A_{11}B_{11} + A_{11}B_{22} + A_{22}B_{11} + A_{22}B_{22} - A_{22}B_{11} + A_{22}B_{21} - A_{11}B_{22} - A_{12}B_{22} + A_{12}B_{21} + A_{12}B_{22} - A_{22}B_{21} - A_{22}B_{22} \\
C_{11} &= A_{11}B_{11} + A_{12}B_{21}
\end{aligned}
$$



I
II
III
IV
V
VI
VII
$C_{11} =$
$A_{11}B_{11}$ $A_{12}B_{11}$ $A_{21}B_{11}$ $A_{22}B_{11}$
$A_{11}B_{21}$ $A_{12}B_{21}$ $A_{21}B_{21}$ $A_{22}B_{21}$
$A_{11}B_{11}$ $A_{12}B_{12}$ $A_{21}B_{12}$ $A_{22}B_{12}$
$A_{11}B_{22}$ $A_{12}B_{22}$ $A_{21}B_{22}$ $A_{22}B_{22}$
$C_{12} =$
$A_{11}B_{11}$ $A_{12}B_{11}$ $A_{21}B_{11}$ $A_{22}B_{11}$
$A_{11}B_{21}$ $A_{12}B_{21}$ $A_{21}B_{21}$ $A_{22}B_{21}$
$A_{11}B_{11}$ $A_{12}B_{12}$ $A_{21}B_{12}$ $A_{22}B_{12}$
$A_{11}B_{22}$ $A_{12}B_{22}$ $A_{21}B_{22}$ $A_{22}B_{22}$
$C_{21} =$
$A_{11}B_{11}$ $A_{12}B_{11}$ $A_{21}B_{11}$ $A_{22}B_{11}$
$A_{11}B_{21}$ $A_{12}B_{21}$ $A_{21}B_{21}$ $A_{22}B_{21}$
$A_{11}B_{11}$ $A_{12}B_{12}$ $A_{21}B_{12}$ $A_{22}B_{12}$
$A_{11}B_{22}$ $A_{12}B_{22}$ $A_{21}B_{22}$ $A_{22}B_{22}$
$C_{22} =$
$A_{11}B_{11}$ $A_{12}B_{11}$ $A_{21}B_{11}$ $A_{22}B_{11}$
$A_{11}B_{21}$ $A_{12}B_{21}$ $A_{21}B_{21}$ $A_{22}B_{21}$
$A_{11}B_{11}$ $A_{12}B_{12}$ $A_{21}B_{12}$ $A_{22}B_{12}$
$A_{11}B_{22}$ $A_{12}B_{22}$ $A_{21}B_{22}$ $A_{22}B_{22}$

Strassen's Algorithm

$$
\begin{aligned}
\mathrm{I} &= (A_{11} + A_{22}) \cdot (B_{11} + B_{22}) \\
\mathrm{II} &= (A_{21} + A_{22}) \cdot B_{11} \\
\mathrm{III} &= A_{11} \cdot (B_{12} - B_{22}) \\
\mathrm{IV} &= A_{22} \cdot (-B_{11} + B_{21}) \\
\mathrm{V} &= (A_{11} + A_{12}) \cdot B_{22} \\
\mathrm{VI} &= (-A_{11} + A_{21}) \cdot (B_{11} + B_{12}) \\
\mathrm{VII} &= (A_{12} - A_{22}) \cdot (B_{21} + B_{22})
\end{aligned}
$$

$$
\begin{aligned}
C_{11} &= \mathrm{I} + \mathrm{IV} - \mathrm{V} + \mathrm{VII} \\
C_{21} &= \mathrm{II} + \mathrm{IV} \\
C_{12} &= \mathrm{III} + \mathrm{V} \\
C_{22} &= \mathrm{I} + \mathrm{III} - \mathrm{II} + \mathrm{VI}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{I} &= (\mathbf{A}_{11} + \mathbf{A}_{22}) \cdot (\mathbf{B}_{11} + \mathbf{B}_{22}) \\
\mathbf{II} &= (\mathbf{A}_{21} + \mathbf{A}_{22}) \cdot \mathbf{B}_{11} \\
\mathbf{III} &= \mathbf{A}_{11} \cdot (\mathbf{B}_{12} - \mathbf{B}_{22}) \\
\mathbf{IV} &= \mathbf{A}_{22} \cdot (-\mathbf{B}_{11} + \mathbf{B}_{21}) \\
\mathbf{V} &= (\mathbf{A}_{11} + \mathbf{A}_{12}) \cdot \mathbf{B}_{22} \\
\mathbf{VI} &= (-\mathbf{A}_{11} + \mathbf{A}_{21}) \cdot (\mathbf{B}_{11} + \mathbf{B}_{12}) \\
\mathbf{VII} &= (\mathbf{A}_{12} - \mathbf{A}_{22}) \cdot (\mathbf{B}_{21} + \mathbf{B}_{22})
\end{aligned}
$$

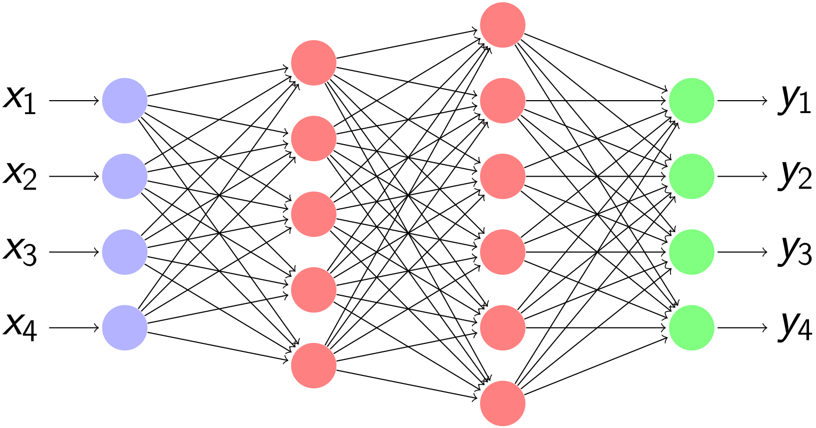
$$
\begin{aligned}
\mathbf{C}_{11} &= \mathbf{I} + \mathbf{IV} - \mathbf{V} + \mathbf{VII} \\
\mathbf{C}_{21} &= \mathbf{II} + \mathbf{IV} \\
\mathbf{C}_{12} &= \mathbf{III} + \mathbf{V} \\
\mathbf{C}_{22} &= \mathbf{I} + \mathbf{III} - \mathbf{II} + \mathbf{VI}
\end{aligned}
$$

Algorithm

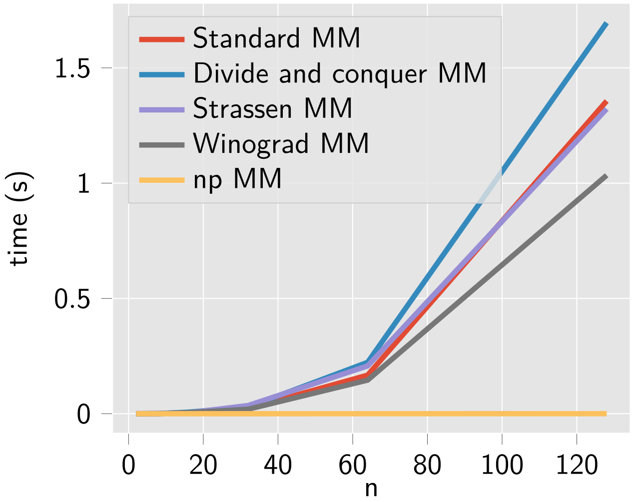
Algorithm 10 Strassen Matrix Multiplication

function $\mathrm{MM}(\mathbf{A}, \mathbf{B}, n)$
    if $n = 2$ **then**
        $\mathbf{C} \leftarrow zeros((n, n))$
        $P \leftarrow (A[0][0] + A[1][1]) \cdot (B[0][0] + B[1][1])$
        $Q \leftarrow (A[1][0] + A[1][1]) \cdot B[0][0]$
        $R \leftarrow A[0][0] \cdot (B[0][1] - B[1][1])$
        $S \leftarrow A[1][1] \cdot (B[1][0] - B[0][0])$
        $T \leftarrow (A[0][0] + A[0][1]) \cdot B[1][1]$
        $U \leftarrow (A[1][0] - A[0][0]) \cdot (B[0][0] + B[0][1])$
        $V \leftarrow (A[0][1] - A[1][1]) \cdot (B[1][0] + B[1][1])$
        $C[0][0] \leftarrow P + S - T + V$
        $C[0][1] \leftarrow R + T$
        $C[1][0] \leftarrow Q + S$
        $C[1][1] \leftarrow P + R - Q + U$
    else
        $m \leftarrow n/2$
        $\mathbf{A11}, \mathbf{A12}, \mathbf{A21}, \mathbf{A22} \leftarrow \mathbf{A}[:m][:m], \mathbf{A}[:m][m:], \mathbf{A}[m:][:m], \mathbf{A}[m:][m:]$
        $\mathbf{B11}, \mathbf{B12}, \mathbf{B21}, \mathbf{B22} \leftarrow \mathbf{B}[:m][:m], \mathbf{B}[:m][m:], \mathbf{B}[m:][:m], \mathbf{B}[m:][m:]$
        $\mathbf{P} \leftarrow \text{strassen}((\mathbf{A11} + \mathbf{A22}), (\mathbf{B11} + \mathbf{B22}), m)$
        $\mathbf{Q} \leftarrow \text{strassen}((\mathbf{A21} + \mathbf{A22}), \mathbf{B11}, m)$
        $\mathbf{R} \leftarrow \text{strassen}(\mathbf{A11}, (\mathbf{B12} - \mathbf{B22}), m)$
        $\mathbf{S} \leftarrow \text{strassen}(\mathbf{A22}, (\mathbf{B21} - \mathbf{B11}), m)$
        $\mathbf{T} \leftarrow \text{strassen}((\mathbf{A11} + \mathbf{A12}), \mathbf{B22}, m)$
        $\mathbf{U} \leftarrow \text{strassen}((\mathbf{A21} - \mathbf{A11}), (\mathbf{B11} + \mathbf{B12}), m)$
        $\mathbf{V} \leftarrow \text{strassen}((\mathbf{A12} - \mathbf{A22}), (\mathbf{B21} + \mathbf{B22}), m)$
        $\mathbf{C11} \leftarrow \mathbf{P} + \mathbf{S} - \mathbf{T} + \mathbf{V}$
        $\mathbf{C12} \leftarrow \mathbf{R} + \mathbf{T}$
        $\mathbf{C21} \leftarrow \mathbf{Q} + \mathbf{S}$
        $\mathbf{C22} \leftarrow \mathbf{P} + \mathbf{R} - \mathbf{Q} + \mathbf{U}$
        $C \leftarrow vstack((hstack((C11, C12)), hstack((C21, C22))))$
    return C

Neural Network



Measurement



Vector-Matrix Multiplication

$$\begin{bmatrix} A_{1,1} & A_{1,2} & A_{1,3} & A_{1,4} \end{bmatrix} \begin{bmatrix} B_{1,1} & B_{1,2} & B_{1,3} & B_{1,4} & B_{1,5} \\ B_{2,1} & B_{2,2} & B_{2,3} & B_{2,4} & B_{2,5} \\ B_{3,1} & B_{3,2} & B_{3,3} & B_{3,4} & B_{3,5} \\ B_{4,1} & B_{4,2} & B_{4,3} & B_{4,4} & B_{4,5} \end{bmatrix}$$

$$\begin{bmatrix} C_{1,1} & C_{1,2} & C_{1,3} & C_{1,4} & C_{1,5} \end{bmatrix}$$

Vector-Matrix Multiplication

$$\begin{bmatrix} A_{1,1} & A_{1,2} & A_{1,3} & A_{1,4} \end{bmatrix} \begin{bmatrix} B_{1,1} & B_{1,2} & B_{1,3} & B_{1,4} & B_{1,5} \\ B_{2,1} & B_{2,2} & B_{2,3} & B_{2,4} & B_{2,5} \\ B_{3,1} & B_{3,2} & B_{3,3} & B_{3,4} & B_{3,5} \\ B_{4,1} & B_{4,2} & B_{4,3} & B_{4,4} & B_{4,5} \end{bmatrix}$$

$$\begin{bmatrix} C_{1,1} & C_{1,2} & C_{1,3} & C_{1,4} & C_{1,5} \end{bmatrix}$$

Vector-Matrix Multiplication

$$\begin{bmatrix} A_{1,1} & A_{1,2} & A_{1,3} & A_{1,4} \end{bmatrix} \begin{bmatrix} B_{1,1} & B_{1,2} & B_{1,3} & B_{1,4} & B_{1,5} \\ B_{2,1} & B_{2,2} & B_{2,3} & B_{2,4} & B_{2,5} \\ B_{3,1} & B_{3,2} & B_{3,3} & B_{3,4} & B_{3,5} \\ B_{4,1} & B_{4,2} & B_{4,3} & B_{4,4} & B_{4,5} \end{bmatrix}$$

$$\begin{bmatrix} C_{1,1} & C_{1,2} & C_{1,3} & C_{1,4} & C_{1,5} \end{bmatrix}$$

Vector-Matrix Multiplication

$$\begin{bmatrix} A_{1,1} & A_{1,2} & A_{1,3} & A_{1,4} \end{bmatrix} \begin{bmatrix} B_{1,1} & B_{1,2} & B_{1,3} & B_{1,4} & B_{1,5} \\ B_{2,1} & B_{2,2} & B_{2,3} & B_{2,4} & B_{2,5} \\ B_{3,1} & B_{3,2} & B_{3,3} & B_{3,4} & B_{3,5} \\ B_{4,1} & B_{4,2} & B_{4,3} & B_{4,4} & B_{4,5} \end{bmatrix}$$

$$\begin{bmatrix} C_{1,1} & C_{1,2} & C_{1,3} & C_{1,4} & C_{1,5} \end{bmatrix}$$

Vector-Matrix Multiplication

$$\begin{bmatrix} A_{1,1} & A_{1,2} & A_{1,3} & A_{1,4} \end{bmatrix} \begin{bmatrix} B_{1,1} & B_{1,2} & B_{1,3} & B_{1,4} & B_{1,5} \\ B_{2,1} & B_{2,2} & B_{2,3} & B_{2,4} & B_{2,5} \\ B_{3,1} & B_{3,2} & B_{3,3} & B_{3,4} & B_{3,5} \\ B_{4,1} & B_{4,2} & B_{4,3} & B_{4,4} & B_{4,5} \end{bmatrix}$$

$$\begin{bmatrix} C_{1,1} & C_{1,2} & C_{1,3} & C_{1,4} & C_{1,5} \end{bmatrix}$$

DSP Architecture