# EE 789 Algorithmic Design of Digital Systems

# Assignment 2

# Matrix Multiplication



Name: Satvikkumar Patel

Roll number: 23M1117

Program: M.tech (EE5 - Electronic Systems)

Department: Electrical Engineering

Date of Sub: November 17, 2023

### EE 789

## Assignment 2: Matrix Multiplication

#### October 9, 2023

You are given two 16x16 matrices A and B and are expected to design a circuit which multiplies these matrices.

If  $A = [a_{ij}]$  and  $B = [b_{ij}]$   $(0 \le i, j \le 15)$  then the product matrix  $C = A \times B$  has entries  $c_{ij}$  with

$$c_{ij} = \sum_{k=0}^{1} 5a_{ik}b_{kj}$$

A 16x16 matrix can be viewed in row-form or column form. In row-form, we can write  $\underline{\ }$ 

$$A = \begin{pmatrix} rA_0^T \\ rA_1^T \\ \dots \\ rA_15^T \end{pmatrix}$$

where  $rA_0^T$  is the first row and so on. In column form, we can write A as

$$A = (cA_0 \ cA_1 \ \dots \ cA_15)$$

where  $cA_0^T$  is the first column and so on. Similarly for B.

The product matrix C can be built in different ways.

$$c_{ij} = rA_i^T.cB_j^T$$

or (as a sum of rank-1 matrices):

$$C = \sum_{i=0}^{15} cA_i \cdot rB_j^T$$

## 1 Sample code

I will share sample code for the multiplication of two 16x16 matrices. This includes a dot-product based implementation and a test-bench.

## 2 Assignment

You will attempt to implement the mmul routine in various ways. The testbench is unchanged.

- Implement the matrix multiplication by speeding up the dot product implementation used in the sample code (10).
- Parallelize the matrix multiplication by dividing A and B into four 8x8 blocks, doing work on the 8x8 blocks and then combining this work to produce the final result (10).
- Implement the matrix product using the sum of rank-1 matrices approach (10).

In each case, you will have to measure the time required by the mmul routine in your implementation.

#### Sample code:

```
1 $parameter ORDER 16
3 //
4 // observation signals to keep track of progress.
5 //
6 $pipe mmul_I : $uint <8> $signal
7 $pipe mmul_J : $uint <8> $signal
8 $pipe mmul_START: $uint <8> $signal
9 $pipe mmul_END: $uint<8> $signal
11 $storage A B C: $array[ORDER][ORDER] $of $uint < 32 >
13 $module [storeA] $in (I J: $uint<8> wval: $uint<32>) $out () $is
14
    A[I][J] := wval
16
17 $module [storeB] $in (I J: $uint<8> wval: $uint<32>) $out () $is
    B[I][J] := wval
19
20 }
21
22 $module [loadC] $in (I J : $uint < 8 > ) $out (Y : $uint < 32 > ) $is
    Y := C[I][J]
24
25 }
26
27
28 $module [dot_product] $in (I J: $uint<8>) $out (result: $uint<32>) $is
29
    $branchblock[loop] {
30
31
      $dopipeline $depth 31 $fullrate
32
33
        $merge $entry $loopback
           $phi K := $zero<8> $on $entry nK $on $loopback
34
           $phi SUMO := ($bitcast ($uint <32>) 0)
                                                     $on $entry nSUMO $on $loopback
35
           $phi SUM1 := ($bitcast ($uint<32>) 0)
                                                     $on $entry nSUM1 $on $loopback
36
        $endmerge
38
        volatile nK := (K + 2)
        $volatile continue_flag := (K < (ORDER-2))</pre>
39
        volatile K1 := (K+1)
40
41
        nSUMO := (SUMO + (A[I][K] * B[K][J]))
42
        nSUM1 := (SUM0 + (A[I][K1] * B[K1][J]))
43
      $while continue_flag
44
    \} (nSUMO => RO nSUM1 => R1)
45
    $volatile result := (R0 + R1)
46
47 }
48
  $module [mmul] $in () $out () $is
49
50
    mmul_START := 1
51
    $branchblock[loop] {
        u := mmul_START v := mmul_END // for logging
54
        $merge $entry I_loopback
        $phi I := $zero<8> $on $entry nI $on I_loopback
      $endmerge
56
      volatile nI := (I + 1)
57
58
      $dopipeline $depth 3
```

```
60
         $merge $entry $loopback
           $phi J := $zero<8> $on $entry nJ $on $loopback
61
         $endmerge
62
         volatile nJ := (J + 1)
63
         $volatile continue_flag := (J < (ORDER - 1))</pre>
64
65
         // dot-product!
66
         C[I][J] := ($call dot_product (I J))
67
         mmul_I := I
69
         mmul_J := J
70
71
      $while continue_flag
72
73
         $if (I < (ORDER-1)) $then $place [I_loopback] $endif</pre>
74
    }
75
    mmul_END := 1
76
77 }
```

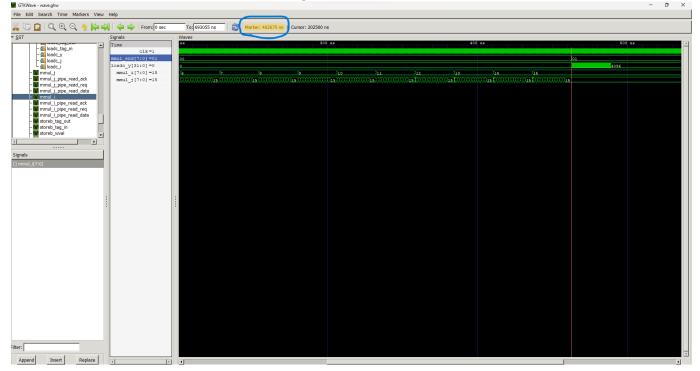
#### TestBench:

```
# #include <signal.h>
2 #include <stdio.h>
3 #include <stdlib.h>
4 #include <stdint.h>
5 #include <pthread.h>
6 #include <pthreadUtils.h>
7 #include <Pipes.h>
8 #include <pipeHandler.h>
9 #ifndef SW
#include "vhdlCStubs.h"
11 #endif
13 #define ORDER 16
14
15 // Give inputs as
16 // Matrix A:
17 // 1
        2
           3 4
                  5
                        7
                            8
                              9 10 11 12 13 14 15 16
                     6
        2
                         7
            3
             4
                  5
                     6
                            8
                               9 10 11 12 13 14 15 16
        2
            3
                  5
                     6
                         7
                            8
                               9 10 11 12 13 14 15 16
19 // 1
               4
                  5
        2
            3
              4
                     6
                         7
                            8
                               9 10 11 12 13 14 15 16
20 // 1
21 //
23 // Matrix B:
24 // 1
       1
           1
               1
                  1
                     1
                                  1
                        1
                            1
                               1
                                     1
                                         1
                                            1
25 // 2
           2
        2
               2
                  2
                     2
                         2
                            2
                               2
                                  2
                                     2
                                         2
                                            2
                                               2
                                                  2 2
26 // 3 3
           3 3
                  3
                     3
                         3
                            3
                               3
                                  3
                                     3
                                         3
                                            3
                                               3
                                                  3
                                                      3
27 // 4 4
           4
                  4
                     4
                         4
                            4
                               4
                                  4
              4
28 //
30
int main(int argc, char* argv[])
32 {
    int I, J;
33
    for (I = 0; I < ORDER; I++)
34
35
      for(J = 0; J < ORDER; J++)
36
37
        storeA(I, J,(uint32_t)
                                  I+1);
38
         storeB(I, J,(uint32_t)
                                  J+1);
39
      }
40
    }
41
```

```
fprintf(stderr, "Stored A, B\n");
42
43
    mmul();
44
45
    fprintf(stderr,"finished dot_product, results:\n");
46
    for (I = 0; I < ORDER; I++)
47
48
       for(J = 0; J < ORDER; J++)
49
50
         uint32_t result = loadC (I,J);
51
         fprintf(stderr, "C[%d][%d] = %d.\n", I, J, result);
52
       }
54
    }
55
56
    return(0);
57
<sub>58</sub> }
```

#### **Results:**

The time taken for the mmul\_end variable to go 1, is 462675ns



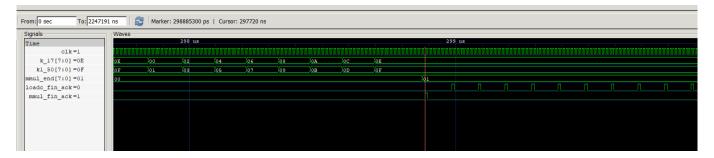
#### Faster Dot product:

The method used here is unrolling the pipe by 2. The matrix multiplication part is left as it is.

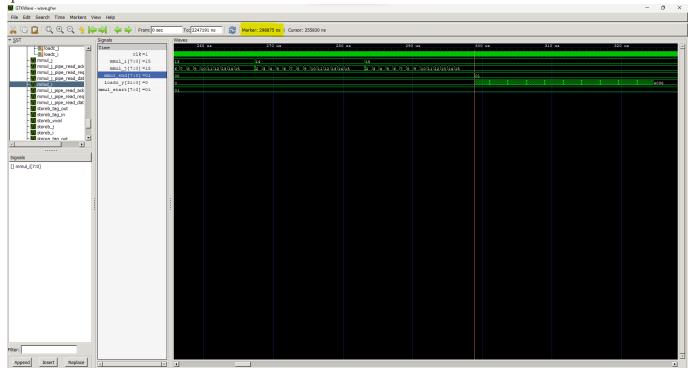
```
1 $parameter ORDER 16
2
3 //
_{4} // observation signals to keep track of progress.
5 //
6 $pipe mmul_I : $uint <8> $signal
7 $pipe mmul_J : $uint < 8 > $signal
8 $pipe mmul_START: $uint <8> $signal
9 $pipe mmul_END: $uint <8> $signal
  $storage A B C: $array[ORDER][ORDER] $of $uint<32>
$module [storeA] $in (I J: $uint<8> wval: $uint<32>) $out () $is
14 {
    A[I][J] := wval
16 }
17 $module [storeB] $in (I J: $uint<8> wval: $uint<32>) $out () $is
    B[I][J] := wval
19
20 }
21
22 $module [loadC] $in (I J : $uint<8>) $out (Y : $uint<32>) $is
23 €
    Y := C[I][J]
24
25
 }
26
27
28 $module [dot_product] $in (I J: $uint < 8 >) $out (result: $uint < 32 >) $is
29
    $branchblock[loop] {
30
31
      $dopipeline $depth 31 $fullrate
32
         $merge $entry $loopback
           $phi K := $zero<8> $on $entry nK $on $loopback
34
           $phi SUMO := ($bitcast ($uint < 32 >) 0)
                                                    $on $entry nSUMO $on $loopback
35
           $phi SUM1 := ($bitcast ($uint<32>) 0)
                                                     $on $entry nSUM1 $on $loopback
36
         $endmerge
37
         volatile nK := (K + 2)
38
         $volatile continue_flag := (K < (ORDER-2))</pre>
39
         volatile K1 := (K+1)
40
        nSUMO := (SUMO + (A[I][K] * B[K][J]))
42
        nSUM1 := (SUM0 + (A[I][K1] * B[K1][J]))
43
      $while continue_flag
44
    \} (nSUMO => RO nSUM1 => R1)
45
    $volatile result := (R0 + R1)
46
47 }
48
  $module [mmul] $in () $out () $is
49
50
    mmul_START := 1
51
    $branchblock[loop] {
52
        u := mmul_START v := mmul_END // for logging
53
        $merge $entry I_loopback
54
         phi I := \frac{2e}{6} $on $entry nI $on I_loopback
56
      $endmerge
57
      volatile nI := (I + 1)
58
```

```
$dopipeline $depth 3
59
         $merge $entry $loopback
           phi J := \frac{8}{6}  son entry nJ  son nJ 
61
         $endmerge
62
         volatile nJ := (J + 1)
63
         $volatile continue_flag := (J < (ORDER - 1))</pre>
64
65
         // dot-product!
66
        C[I][J] := ($call dot_product (I J))
68
        mmul_I := I
69
        mmul_J := J
70
71
      $while continue_flag
72
73
         if (I < (ORDER-1))  then place [I_loopback]  sendif
74
    }
75
76
    mmul_END := 1
77 }
```

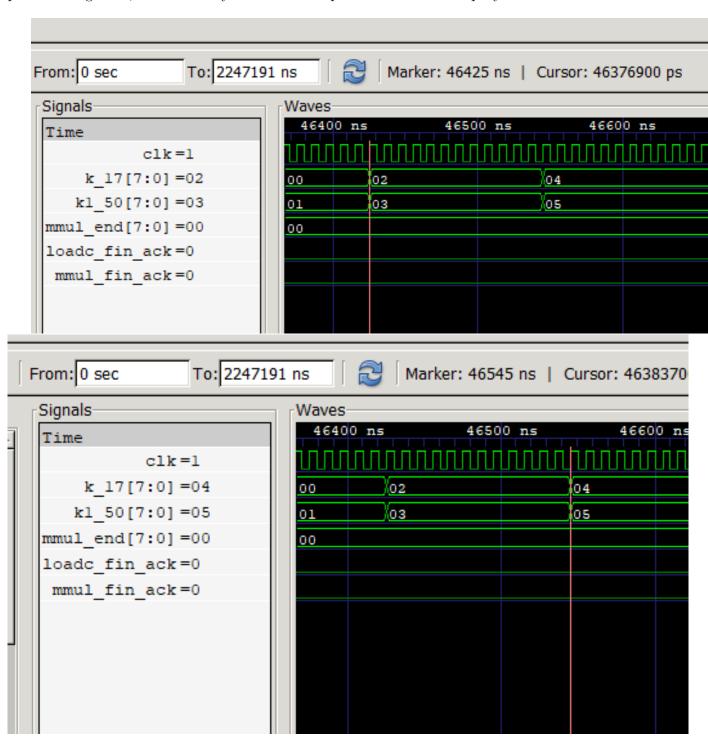
#### **Results:**



The time taken for the mmul\_end variable to go 1, is 298875ns, which is lesser compared to sample code mmul\_end.



The time taken for the loop variable to count up by 1 is, 46545 - 46425 = 120ns. The clock time period being 10ns, it takes 12 cycles for the loop interval to count up by 1.



### Matrix Multiplication by dividing the matrix

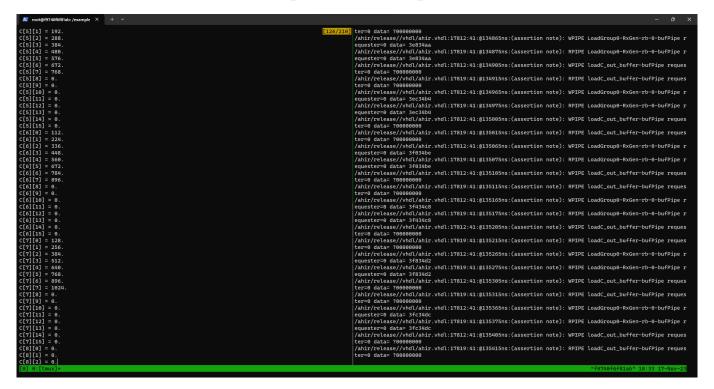
```
1 $parameter ORDER 16
2 $parameter ORDER_by_2 8
5 // observation signals to keep track of progress.
6 //
7 $pipe mmul_I : $uint <8> $signal
8 $pipe mmul_J : $uint < 8 > $signal
9 $pipe mmul_START: $uint <8> $signal
10 $pipe mmul_END: $uint<8> $signal
12 $storage A B C: $array[ORDER][ORDER] $of $uint < 32 >
14 $module [storeA] $in (I J: $uint<8> wval: $uint<32>) $out () $is
    A[I][J] := wval
16
17 }
18 $module [storeB] $in (I J: $uint<8> wval: $uint<32>) $out () $is
    B[I][J] := wval
20
  }
21
23 $module [loadC] $in (I J : $uint < 8 > ) $out (Y : $uint < 32 > ) $is
24
    Y := C[I][J]
25
26 }
27
  $module [dot_product] $in (I J: $uint < 8 >) $out (result: $uint < 32 >) $is
30
    $branchblock[loop] {
31
33
      $dopipeline $depth 31 $fullrate
        $merge $entry $loopback
           hi K := \frac{8}{2}  son hi K := \frac{8}{2}
35
           $phi SUMO := ($bitcast ($uint < 32 >) 0)
                                                     $on $entry nSUMO $on $loopback
36
           $phi SUM1 := ($bitcast ($uint<32>) 0)
                                                     $on $entry nSUM1 $on $loopback
38
        $endmerge
        volatile nK := (K + 2)
39
        $volatile continue_flag := (K < (ORDER-2))</pre>
40
        volatile K1 := (K+1)
41
42
        nSUMO := (SUMO + (A[I][K] * B[K][J]))
43
        nSUM1 := (SUM0 + (A[I][K1] * B[K1][J]))
44
      $while continue_flag
    \} (nSUMO => RO nSUM1 => R1)
46
    $volatile result := (R0 + R1)
47
48
49
50 $module [mmul] $in () $out () $is
  {
51
    mmul_START := 1
    $branchblock[loop1] {
        u := mmul_START v := mmul_END // for logging
54
        $merge $entry I_loopback
        $phi I := $zero<8> $on $entry nI $on I_loopback
56
      $endmerge
57
      volatile nI := (I + 1)
58
```

```
$dopipeline $depth 3
60
         $merge $entry $loopback
           phi J := \frac{8}{6} son entry nJ son sloopback
         $endmerge
63
         volatile nJ := (J + 1)
64
         $volatile continue_flag := (J < (ORDER_by_2 - 1))</pre>
65
         // dot-product!
67
         C[I][J] := ($call dot_product (I J))
         mmul_I := I
         mmul_J := J
71
72
       $while continue_flag
73
74
         $if (I < (ORDER_by_2-1)) $then $place [I_loopback] $endif
     }
76
77
78
     $branchblock[loop2] {
79
         u := mmul_START v := mmul_END // for logging
80
81
         $merge $entry I_loopback
         phi I := \frac{8}{6} $on $entry nI $on I_loopback
82
       $endmerge
       volatile nI := (I + 1)
85
       $dopipeline $depth 3
86
87
         $merge $entry $loopback
           $phi J := $zero<8> $on $entry nJ $on $loopback
         $endmerge
89
         volatile nJ := (J + 1)
90
         $volatile continue_flag := (J < (ORDER_by_2 - 1))</pre>
         volatile I_1 := (I + 8)
92
         volatile J_1 := (J + 8)
93
94
95
         // dot-product!
         C[I_1][J_1] := (call dot_product (I_1 J_1))
96
97
         mmul_I := I
98
         mmul_J := J
100
       $while continue_flag
         $if (I < (ORDER_by_2-1))    $then    $place [I_loopback]    $endif
103
     }
104
105
     $branchblock[loop3] {
106
         u := mmul_START v := mmul_END // for logging
         $merge $entry I_loopback
108
         phi I := \frac{8}{6}
109
       $endmerge
110
       volatile nI := (I + 1)
111
112
       $dopipeline $depth 3
113
         $merge $entry $loopback
114
           $phi J := $zero<8> $on $entry nJ $on $loopback
         $endmerge
116
         volatile nJ := (J + 1)
117
         $volatile continue_flag := (J < (ORDER_by_2 - 1))</pre>
118
119
         //$volatile I_1 := (I + 8)
         volatile J_1 := (J + 8)
```

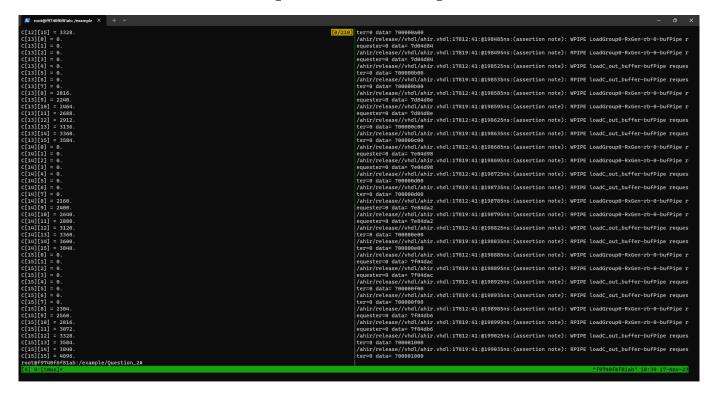
```
// dot-product!
         C[I][J_1] := (call dot_product (I J_1))
123
124
         mmul_I := I
         mmul_J := J
126
127
       $while continue_flag
128
         $if (I < (ORDER_by_2-1)) $then $place [I_loopback] $endif</pre>
130
131
132
     $branchblock[loop4] {
133
         u := mmul_START v := mmul_END // for logging
134
         $merge $entry I_loopback
135
         phi I := \frac{8}{6} on prime I := \frac{8}{6}
136
       $endmerge
137
       volatile nI := (I + 1)
138
139
       $dopipeline $depth 3
140
         $merge $entry $loopback
141
            phi J := \frac{8}{6} son entry nJ son sloopback
142
          $endmerge
143
         volatile nJ := (J + 1)
144
          $volatile continue_flag := (J < (ORDER_by_2 - 1))</pre>
145
146
          volatile I_1 := (I + 8)
         //$volatile J_1 := (J + 8)
147
148
         // dot-product!
149
         C[I_1][J] := (call dot_product (I_1 J))
150
151
         mmul_I := I
         mmul_J := J
153
154
       $while continue_flag
155
156
          $if (I < (ORDER_by_2-1)) $then $place [I_loopback] $endif</pre>
157
     }
158
159
     mmul_END := 1
160
161 }
```

Four branch blocks are used, in order to find the resultant multiplication matrix. - The loop 1 is finding the top left matrix.

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{18} \\ a_{21} & a_{22} & \cdots & a_{28} \\ \vdots & \vdots & \ddots & \vdots \\ a_{81} & a_{82} & \cdots & a_{88} \end{bmatrix}$$



- The loop 2 is finding the bottom right matrix.



- The loop 3 is finding bottom left matrix.

```
\begin{bmatrix} a_{911} & a_{912} & \cdots & a_{98} \\ a_{1011} & a_{1012} & \cdots & a_{108} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1611} & a_{1612} & \cdots & a_{168} \end{bmatrix}
```

```
| Color | Colo
```

23M1117 14 November 17, 2023

- The loop 4 is finding top right matrix.

```
\begin{bmatrix} a_{19} & a_{110} & \cdots & a_{116} \\ a_{29} & a_{210} & \cdots & a_{216} \\ \vdots & \vdots & \ddots & \vdots \\ a_{89} & a_{810} & \cdots & a_{816} \end{bmatrix}
```

```
root@f9740f6f81ab: /example X
C[12][15] = 3328.
C[13][0] = 224.
C[13][1] = 448.
C[13][2] = 672.
C[13][3] = 896.
C[13][4] = 1120.
C[13][5] = 1344.
C[13][6] = 1568.
C[13][7] = 1792.
C[13][8] = 2016.
C[13][9] = 2240.
C[13][10] = 2464.
C[13][11] = 2688.
C[13][12] = 2912.
C[13][13] = 3136.
C[13][14] = 3360.
C[13][15] = 3584.
C[14][0] = 240.
C[14][1] = 480.

C[14][2] = 720.
C[14][3] = 960.
C[14][4] = 1200.
C[14][5] = 1440.
C[14][6] = 1680.
C[14][7] = 1920.
C[14][8] = 2160.
C[14][9] = 2400.
C[14][10] = 2640.
C[14][11] = 2880.
C[14][12] = 3120.
C[14][13] = 3360.

C[14][14] = 3600.
C[14][15] = 3840.
C[15][\theta] = 256.
C[15][1] = 512.
C[15][2] = 768.
C[15][3] = 1024.
C[15][4] = 1280.
C[15][5] = 1536.
C[15][6] = 1792.
C[15][7] = 2048.

C[15][8] = 2304.
C[15][9] = 2560.
C[15][10] = 2816.
C[15][11] = 3072.
C[15][12] = 3328.
C[15][13] = 3584.
C[15][14] = 3840.
C[15][15] = 4096.
root@f9740f6f81ab:/example/Question_2#
```

#### Results:

The time difference in finishing the matrix multiplication is can be seen. The right side waveforms are of the sample code while the left side waveforms is the modified matrix multiplication algorithm. The sample code mmul\_end is being reported at 462675ns and the modified code mmul\_end is reported at 300625ns.

