



ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ

Σχεδίαση Συστημάτων Υλικού - Λογισμικού

Εργαστήριο 1: Εξοικείωση με το εργαλείο Vitis
HLS

Από τους φοιτητές:
Σωτήριος Καραντζούλης 10952
Κωνσταντίνος Γκαρίπης 10924

Διδάσκων: Ιωάννης Παπαευσταθίου
Υπεύθυνοι Εργαστηρίου: Α. Αθανασιάδης - Δ. Καρανάσσος

November 28, 2025

Περιεχόμενα

1 Ερώτημα 1: Κώδικας Επιταχυντή και Testbench	3
1.1 Περιγραφή Λειτουργικότητας	3
1.2 Source Code (Accelerator)	3
1.3 Testbench Code	4
2 Ερώτημα 2: Αποτελέσματα Αρχικής Σύνθεσης (C Synthesis)	6
3 Ερώτημα 3: Αποτελέσματα C/RTL Cosimulation	6
4 Ερώτημα 4: Βελτιστοποίηση με Vitis HLS Directives	7
4.1 (i) Επίδραση Μεγέθους Εικόνας (Scaling)	7
4.2 (ii) Βέλτιστη Ύλοποίηση	7
4.3 (iii) Υπολογισμός Επιτάχυνσης (Speed-up)	8

1 Ερώτημα 1: Κώδικας Επιταχυντή και Testbench

1.1 Περιγραφή Λειτουργικότητας

Σε αυτό το ερώτημα σχεδιάστηκε ο hardware accelerator IMAGE_DIFF_POSTERIZE. Ο επιταχυντής δέχεται δύο εικόνες A και B (πίνακες μεγέθους $HEIGHT \times WIDTH$), υπολογίζει την απόλυτη διαφορά των pixel τους και εφαρμόζει κατωφλίωση (thresholding) με βάση τις τιμές $T1 = 32$ και $T2 = 96$.

1.2 Source Code (Accelerator)

Παρακάτω παρατίθεται ο κώδικας της συνάρτησης IMAGE_DIFF_POSTERIZE:

```
1 #define WIDTH 512
2 #define HEIGHT 512
3 #define T1 32
4 #define T2 96
5 #include <stdint.h>
6
7 void IMAGE_DIFF_POSTERIZE(uint8_t A [HEIGHT][WIDTH], uint8_t B[HEIGHT][WIDTH], uint8_t C[HEIGHT][WIDTH]){
8
9     #pragma HLS INTERFACE mode=bram port=A
10    #pragma HLS INTERFACE mode=bram port=B
11    #pragma HLS INTERFACE mode=bram port=C
12    #pragma HLS INTERFACE s_axilite port=return
13    uint8_t D; // difference of pixel values
14    int16_t temp_d;
15
16    // Partition the arrays so more there can be more access in one
17    // iteration with loop unrolling and pipelining
18    // #pragma HLS ARRAY_PARTITION variable=A type=cyclic factor=64 dim=2
19    // #pragma HLS ARRAY_PARTITION variable=B type=cyclic factor=64 dim=2
20    // #pragma HLS ARRAY_PARTITION variable=C type=cyclic factor=64 dim=2
21
22    // iterate through A and B
23    for(int i = 0 ; i < HEIGHT ; i++){
24
25        // #pragma HLS pipeline
26
27        for(int j = 0 ; j < WIDTH ; j++){
28
29            temp_d = A[i][j] - B[i][j];
30            D = (temp_d < 0)? -temp_d : temp_d; // get the difference of each
31            corresponding pixels
32            if(D < T1){
33                C[i][j] = 0;
34            }
35            else if(D >= T1 && D < T2){
36                C[i][j] = 128;
37            }
38            else{
39                C[i][j] = 255;
40            }
41        }
42    }
43 }
```

```

42 }
43 }
44 }
```

Listing 1: Κώδικας του Accelerator (Source.cpp)

1.3 Testbench Code

Για να γεμίσουμε τους πίνακες A και B επιλέξαμε μία απλή συνάρτηση της γραμμής και στήλης κάθε στοιχείου. Παρακάτω παρατίθεται ο κώδικας του Testbench που χρησιμοποιήθηκε για την επαλήθευση:

```

1 #include <stdio.h>
2 #include <stdint.h>
3
4 #define WIDTH 512
5 #define HEIGHT 512
6
7
8 void IMAGE_DIFF_POSTERIZE(uint8_t A [HEIGHT][WIDTH], uint8_t B[HEIGHT][
    WIDTH], uint8_t C[HEIGHT][WIDTH]);
9
10 int main(){
11     uint8_t A[HEIGHT][WIDTH] ;
12     uint8_t B[HEIGHT][WIDTH] ;
13
14     uint8_t ref_data[HEIGHT][WIDTH] = {};
15     // filling matrices with dummy values
16     for(int i=0; i<HEIGHT; i++){
17
18         for(int j=0; j<WIDTH; j++){
19             A[i][j] = (uint8_t) (i*j) % 256;
20             B[i][j] = (uint8_t) (i/(j+1)) % 256;
21         }
22     }
23
24     uint8_t C[HEIGHT][WIDTH] = {};
25
26     IMAGE_DIFF_POSTERIZE(A, B, C);
27
28     FILE *ref= fopen("C:/Users/user/Desktop/ref_output.dat", "r");
29
30     if(ref==NULL){
31
32         FILE *out = fopen("C:/Users/user/Desktop/ref_output.dat", "w");
33         if (!out) {
34             printf("File open error");
35             return 1;
36         }
37
38         for(int i = 0 ; i < HEIGHT ; i++){
39
40             for(int j =0 ; j < WIDTH ; j++){
41                 fprintf(out, "%3d ", C[i][j]);
42             }
43             fprintf(out, "\n");
44         }
45         fclose(out);
```

```

46
47     printf("Reference output from C simulation written.\n");
48     return 0;
49 }
50
51     for (int i = 0; i < HEIGHT; i++) {
52         for(int j = 0; j< WIDTH; j++ ){
53             if (fscanf(ref, "%3d", &ref_data[i][j]) != 1) {
54                 printf("Error: reference file corrupted at index (%d,%d).\n"
55 , i,j);
56                 fclose(ref);
57                 return 1;
58             }
59         }
60     fclose(ref);
61
62     bool match = 1;
63     for (int i = 0; i < HEIGHT; i++) {
64         for(int j = 0; j< WIDTH; j++ ){
65             if(C[i][j]!= ref_data[i][j]){
66                 printf("Mismatch at index (%d,%d)", i,j);
67                 match = 0;
68             }
69         }
70     }
71     if(match){
72         printf("Test PASSED!\n");
73     }
74     else{
75         printf("Test FAILED!\n");
76     }
77     return 0;
78 }
79 }
```

Listing 2: Κώδικας του Testbench (tb.cpp)

2 Ερώτημα 2: Αποτελέσματα Αρχικής Σύνθεσης (C Synthesis)

Η σύνθεση πραγματοποιήθηκε με τις default ρυθμίσεις του εργαλείου, χωρίς τη χρήση directives βελτιστοποίησης.

Διαστάσεις Πινάκων: HEIGHT = 256, WIDTH = 256.

Τα αποτελέσματα που προέκυψαν από το *Synthesis Report* παρουσιάζονται στον παρακάτω πίνακα:

Metric	Value
Estimated Clock Period	5.093 ns
Worst Case Latency	65538 cycles
Number of DSP48E	0
Number of BRAMs	0
Number of FFs	57
Number of LUTs	272

Σχολιασμός: Παρατηρούμε μικρή αξιοποίηση πόρων κατά την σύνθεση και μεγάλες τιμές στο latency, αποτελέσματα λογικά όταν δεν εφαρμόζουμε κάποια βελτιστοποίηση. Η μηδενική χρήση BRAM ήταν αναμενόμενη παρόλο που δηλώνεται σαν Interface στην υλοποίηση, διότι ωστόσο αποδούσε σαν πόρος εξωτερικά του IP core το οποίο ωστόσο συνθέσει το Vitis. Στην συνέχεια, ωστόσο παρατηρήσουμε την αύξηση των πόρων που χρησιμοποιούμε όσο μειώνουμε το latency.

3 Ερώτημα 3: Αποτελέσματα C/RTL Cosimulation

Πραγματοποιήθηκε C/RTL Cosimulation για να επιβεβαιωθεί η ορθότητα της σχεδίασης σε επίπεδο RTL και να μετρηθεί ο ακριβής χρόνος εκτέλεσης.

Διαστάσεις Πινάκων: HEIGHT = 256, WIDTH = 256.

Πίνακας 1: Αποτελέσματα C/RTL Cosimulation

Metric	Value
Total Execution Time	655595 ns
Min Latency	65536 cycles
Avg Latency	65536 cycles
Max Latency	65536 cycles

4 Ερώτημα 4: Βελτιστοποίηση με Vitis HLS Directives

4.1 (i) Επίδραση Μεγέθους Εικόνας (Scaling)

Κρατώντας σταθερό το HEIGHT = 256 και μεταβάλλοντας το WIDTH, παρατηρήθηκαν τα εξής σχετικά με το Latency και τον χρόνο εκτέλεσης:

Πίνακας 2: Επίδραση μεταβολής πλάτους (WIDTH) με σταθερό HEIGHT

HEIGHT	WIDTH	Total Latency (cycles)	Execution Time
256	64	256	2975
256	256	512	5355
256	512	1024	10655

Παρατηρήσεις: Παρατηρούμε ότι μεταβάλλοντας το WIDTH για σταθερό HEIGHT = 256 το execution time και το latency παραμένουν αυξάνονται.

4.2 (ii) Βέλτιστη Υλοποίηση

Μετά από πειραματισμό με διάφορα directives (PIPELINE, UNROLL, ARRAY_PARTITION), η βέλτιστη υλοποίηση για διαστάσεις 256×256 δίνει τα παρακάτω αποτελέσματα:

Directives που χρησιμοποιήθηκαν:

- `#pragma HLS ARRAY_PARTITION variable=A type=cyclic factor=64 dim=2` για όλους τους πίνακες
- `#pragma HLS PIPELINE II=1` στο εσωτερικό loop

Πίνακας 3: Σύγκριση Πόρων και Επίδοσης Βέλτιστης Υλοποίησης

Metric	Value (Optimized)
Estimated Clock Period	4.939 ns
DSP48E Used	0
BRAMs Used	0
FFs Used	58
LUTs Used	26497
Timing (Cosimulation)	
Total Execution Time	5355 ns
Min Latency	512
Avg Latency	512
Max Latency	512

Αιτιολόγηση: Το ARRAY_PARTITION χρησιμοποιήθηκε για να χωριστούν οι πίνακες σε περισσότερα memory banks ώστε να επιτρέπεται η ταυτόχρονη ανάγνωση πολλών δεδομένων. Ο τύπος cyclic επιλέχθηκε ως πλέον κατάλληλος για column traversal όπως εδώ

(εξού και dim=2). Στο factor δόθηκε η τιμή 64 έπειτα από μία διαδικασία trial and error ώστε να βρεθεί το μέγιστο factor που δεν παράγει error στο cosimulation για 256x256 Πίνακες.

Το Pipeline χρησιμοποιήθηκε ως το κύριο directive που επιτυγχάνει το speedup. Με αυτόν τον τρόπο, στο εσωτερικό loop οι τιμές των πινάκων διαβάζονται με διοχέτευση χωρίς δηλαδή να περιμένει το επόμενο iteration το προηγόυμενο. Άλλωστε, κάθε iteration είναι εντελώς ανεξάρτητο από όλα τα υπόλοιπα.

Σημείωση: Επιχειρήθηκε, επίσης, να χρησιμοποιηθεί το directive #pragma HLS unroll το οποίο υλοποιεί loop unrolling στο εσωτερικό loop, χωρίς κάποια ιδιαίτερη διαφορά. Η ιδέα ήταν να εκτελούνται κάποιες επαναλήψεις ταυτόχρονα με loop unrolling και να αρχίζουν και οι επόμενες με pipeling. Δυστυχώς, ο συνδιασμός αυτός απορρίφθηκε χάρην απλότητος στον κώδικα, εφόσον δεν παρουσιάζόταν διαφορά στο execution time.

4.3 (iii) Υπολογισμός Επιτάχυνσης (Speed-up)

Η επιτάχυνση (Speed-up) που επιτεύχθηκε σε σχέση με την αρχική (non-optimized) υλοποίηση είναι:

$$Speedup = \frac{T_{initial}}{T_{optimized}} = \frac{655595}{5355} = \mathbf{122.43} \quad (1)$$