DAT480: Reconfigurable Computing
Exam - April 2022

Date: April 10, 2021

Time: **14:00-18:00**

Examiner: Ioannis Sourdis

Department: Computer Science and Engineering

Inquiries: contact through phone, phone extension 1744

Duration: 4 hours

Grading scale: 100 points in total

Chalmers:
0: 0%-49%, 3: 50%-64%, 4: 65%-84%, 5: 85%-100%
GU:
Fail (U): 0%-49%, Pass (G): 50%-79%, Pass with Distinction (VG): 80%-100%

Available references: Lecture notes are allowed.
Textbooks or Calculators are not allowed.

General: Submit your solutions, in English, on blank paper sheets. Write legibly; feel free to use figures to get your point across.

The order of answering the questions does not matter (start with the easiest ones).

Please start the solutions for each problem on a new sheet. Please number the sheets so that the solutions are in numerical order.

Note that it is possible to receive partial credit for an answer even if it is not 100% correct.

Your personal identity code is required on each submitted sheet!

Good luck!

Note: *These are example answers to the exam questions for one possible set of randomized variables or explanations as of how a solution to an exam question can be derived.*

Question 1 Reconfigurable Devices: (12,5 points)
Consider the following Boolean function:

F(x1, x2, x3, x4)= A*x1*x2 + Bx3*x4 + C*x1*x4

and map it to the following 3 FPGA devices:

- one with (i) 2-input LUTs,
- one with (ii) 3-input LUTs, and
- one with 4-input LUTs.

What is the delay of the function mapped to each of the above devices considering that:

- The delay of a 2-input AND/OR gate is equal to 1 ns.
- the propagation delay of an SRAM cell zero (0 ns)
- the delay added by wires is (0 ns)

Answer
Similar to lecture on Reconfigurable Devices, slide 20. Need to consider that different sizes of LUTs use different multiplexers.

Question 2 Compute Models: (12,5 points)
Describe the main characteristics of the dynamic streaming dataflow compute model and the basic modules used to implement it.
Described the variant of dynamic streaming dataflow with peeks; what is the extra module needed in the compute model and how does it affect determinism?

Answer
See lecture on Compute Models, slides 16-22

Question 3 Reconfigurable System Architectures: (12,5 points)
What are the three levels of integrating reconfigurable hardware in a traditional system? What differences are there?

Answer
See lecture Reconfigurable System Architectures, slide about "Integration to a traditional computing system".

Differences tradeoffs:

- ways of exchanging data (register file, shared cache hierarchy, shared memory/IO, message passing) and associated communication latency/throughput.
- hardware resources (fewer the closest you get to the CPU)

- modularity / ease of integration. More changes the closest you get to the CPU

Question 4 Application Development in Reconfigurable hardware: (12,5 points)
What are the characteristics of pattern matching on network packet payload (Deep Packet Inspection) that make it suitable for reconfigurable acceleration? Are there any characteristics that don't fit well to reconfigurable hardware?

Answer
See lecture Application Development in Reconfigurable hardware, slide 25

- negative: updating ruleset involves reimplementing the entire design.

Question 5 Software for Spatial computing: (12,5 points)

Consider the following code:

```
for (count=0; ; count += 1) {
  int d = input[count];
  int shift = 0;
  while (d != 0 && ((d & 0x3FF) != 0x291)) {
   shift = shift + 1;
   d = d >> 1;
  }
  output[count] = shift;
}
```

- How can one unroll this variable length loop?
- How can you choose the degree of unrolling?
    - What happens if you select a degree of unrolling much higher than the number of executed iterations?
    - What happens if you select a degree of unrolling much lower than the number of executed iterations?

Answer
Similar to lecture on HLD programming examples in slides 38-39

- degree of unrolling much higher than the number of executed iterations: waste resources and add unnecessary latency to the design.
- degree of unrolling much lower than the number of executed iterations: provided resources is not a problem, one would loose performance (processing throughput) compared to unrolling the loop further.

Question 6 Reconfiguration Management: (12,5 points)
What are the benefits of dynamic partial reconfiguration and how is it supported by modern FPGA devices?

Answer
See lecture on reconfigurable management, slides 13-15

Question 7 Energy Efficiency: (12,5 points)
According to Mike Flynn, an FPGA device (originally created an "emulation" technology to prototype ASICs) can deliver better performance and better performance/Watt for some applications compared to general purpose computers, although its Area – Power - Delay product is 2-3 orders of magnitude worse than ASICs. How is this paradox justified?

Answer
In a general purpose processor chip, only a small fraction of resources is used to carry actual computations (functional units). The rest of the chip resources are memory blocks (caches) or other modules that improve processing (e.g. branch prediction). On the contrary, an FPGA chip can utilize all its resources to perform useful computations. Consequently, although the FPGA A*P*D is 10^2-10^3 worse than ASICs, it can deliver more efficient computations *if* the application characteristics can be matched by the FPGA substrate. In addition, enabling a wide and deep pipeline of computations minimizes data movement reducing execution times and energy costs.

Question 8 Defect and Fault Tolerance: (12,5 points)
- What are the differences and tradeoffs between the following defect tolerance approaches: (i) Defect Map, (ii) Perfect component.
- Which of the two approaches would you choose for making a defect tolerant FPGA device and why?

Answer
See lecture on defect and fault tolerance, slide 13
Defect map would need the synthesis-implementation tools to be made aware of the defects. As opposed to the perfect component which requires no changes in the tools. Although more resource wasteful the second choice would 4probably be preferable because FPGA resources are rich and the defect rates would be quite low to not disable a large fraction of the chip.