

CEG 5010: Cost

Philip Leong (phwl)

*"An engineer is a man who can do for a dime
what any fool can do for a dollar"*

Optimizing for speed

- Most attention has been given to optimizing for speed
 - Both in literature and in this course
- Many systems where this is not the best metric
 - e.g. signal processing systems where the sample rate is fixed
 - Audio signal processing has normally 2×44.1 KHz inputs

Some of the design styles we have studied

- Parallel
- Bit-serial
- Digit serial
- Systolic
- Pipelining

Each give a different point on the area/time curve.

Bigger picture:

so do full custom VLSI, gate arrays, FPGAs, microprocessors

Cost

- When we make a design we need to consider what we are trying to optimize
 - Area
 - Speed
 - Throughput
 - Latency
 - Area*time
 - Power
 - Design time
 - Component cost
 - Flexibility
- Obviously this affects our architecture

Area vs time

- Often conflicting
- Consider multiplier
 - Fixed point or floating point
 - Pipelined
 - How parallel?
 - Bit-serial, digit serial, parallel (what about the adders?)
 - Systolic
 - Distributed arithmetic
- When you make a design, consider what can be done in parallel and what makes the best compromise between area and time
- Also think about what proportion of the time resources are idle

Other ideas to consider

Module Generators

- It is possible to make generalized designs which give more than a single point on the area/time chart
- Some examples
 - digit serial multiplier can give different area/time values depending on the digit size
 - Controlled unrolling of loops
for (sum = i = 0; i < 100; i++)
sum = sum + i;
- Not much harder to produce than a single design, but gives much more flexibility in exploring area/time tradeoffs

Multicontext Designs

- All designs presented so far have a single context
- How about reusing resources?
 - FPGA resources used like physical memory in a VM system
 - Can be considered if there are idle resources
 - i.e. Context switching must be much lower than computationally active resources
 - More efficient utilization
 - FPGA resources are loaded on demand so they can be shared
 - At the cost of
 - Scheduling issues
 - Requiring special FPGAs to support this type of architecture
- Question: can you think of situations when it can/cannot be used?
- Question: what changes to the standard FPGA architecture are required?