UNIVERSITÀ DELLA SVIZZERA ITALIANA

Meeting - 23.05.2014

Simon Maurer

May 22, 2014

1 ARCHITECTURE

How does the architecture look like:

- big pipeline to reduce memory access and parallelize computation of likelihood
- pipeline each stage of the big pipeline in order to maximize MACC throughput
- propose two solutions for the stages of the big pipeline:
  - serial multiplication (register to DSP slice ratio is high) -> use ram to save intermediate steps to reduce register usage?
  - parallel multiplication (register to DSP slice ratio is low) -> needs a powerful memory hierarchy to actually make sense (or maybe hardcode the parameters into a rom?)
- use a controller to govern the pipelines

2 SCALING AND PRECISION

Some important pointis about scaling:

- neumann uses division on each element. Better: calculate a scaling factor with division operation and multiply with each element
- does not influence the order of complexity, has however a big impact on necessary recources.

- tradeoff: use scaling (more recources) lower precision needed (less recources) / no scaling (less recources) higher precision needed (more recources)
- if scaling, then use different method than the proposed one:
  - avoid division (use shift right)
  - avoid log function for likelihood (compute likelihood by accumulation and shift left. ATTENTION: overflow)
  - either use average calculation to shift a fixed amount (check for overflow) or use the comparation element of the DSP slice.

and about precision (floating point vs fixed point):

- floating points are complicated
- floating points need more recources
- floating points are error prone
- floating points are precise
- fixed points are less precise (for a given width)
- fixed points are faster
- fixed points are easier to debug

-> use fixed point and simple scaling. Give precision ranges and design the system with higher operant width.

3 TESTING

- use Nexis4 board [1]
- store parameters and bitstream into flash, then load parameters into ram
- check memory solutions on the market to increase the data throughput and increase the speed of the whole system (cf. parallelized pipline stage)
- provide sequence stream either by loading a testsample into flash or by using an external interface (ethernet, serial port, etc)

[1] http://www.xilinx.com/products/boards-and-kits/1-3YZNP5.htm