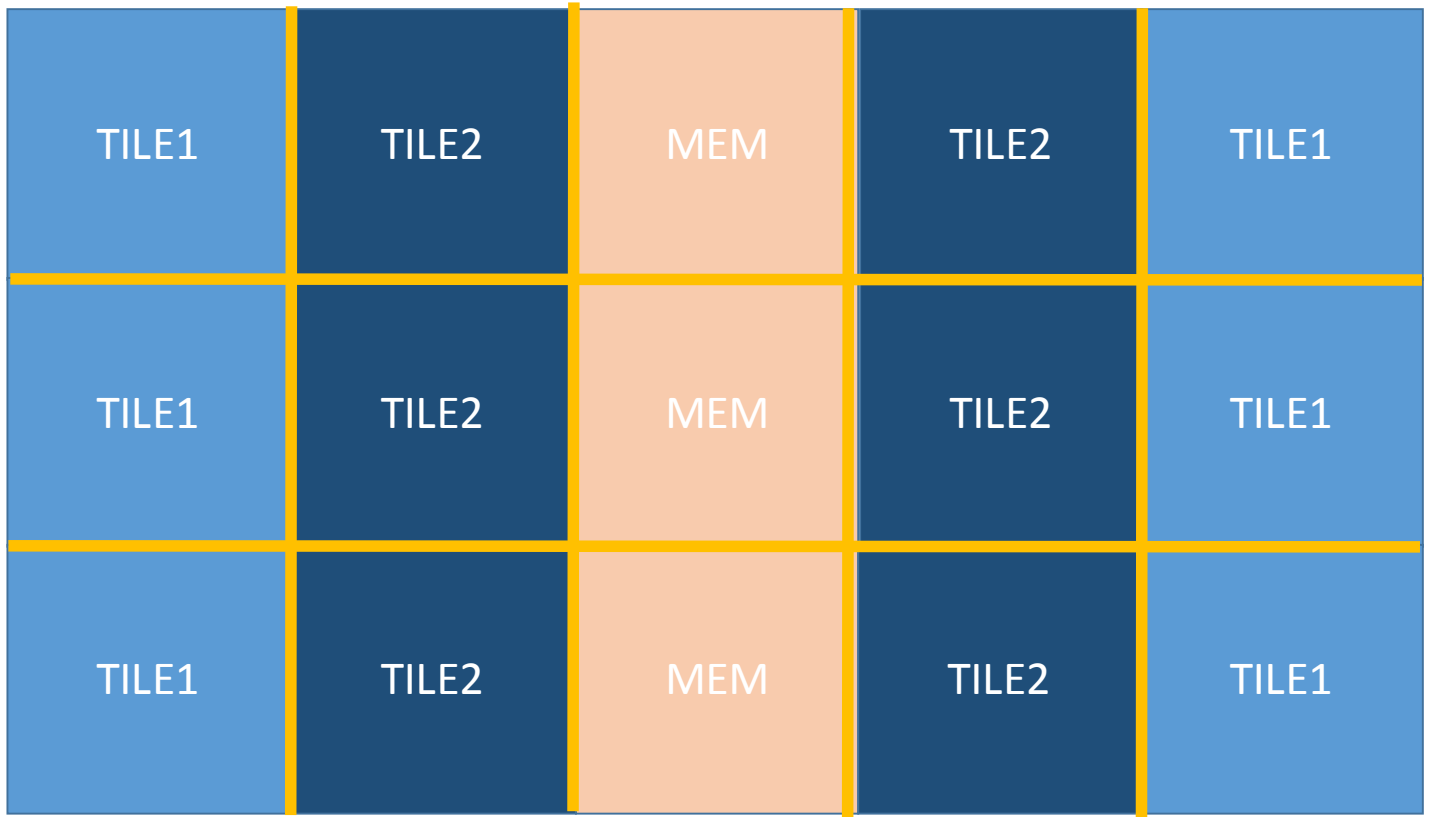
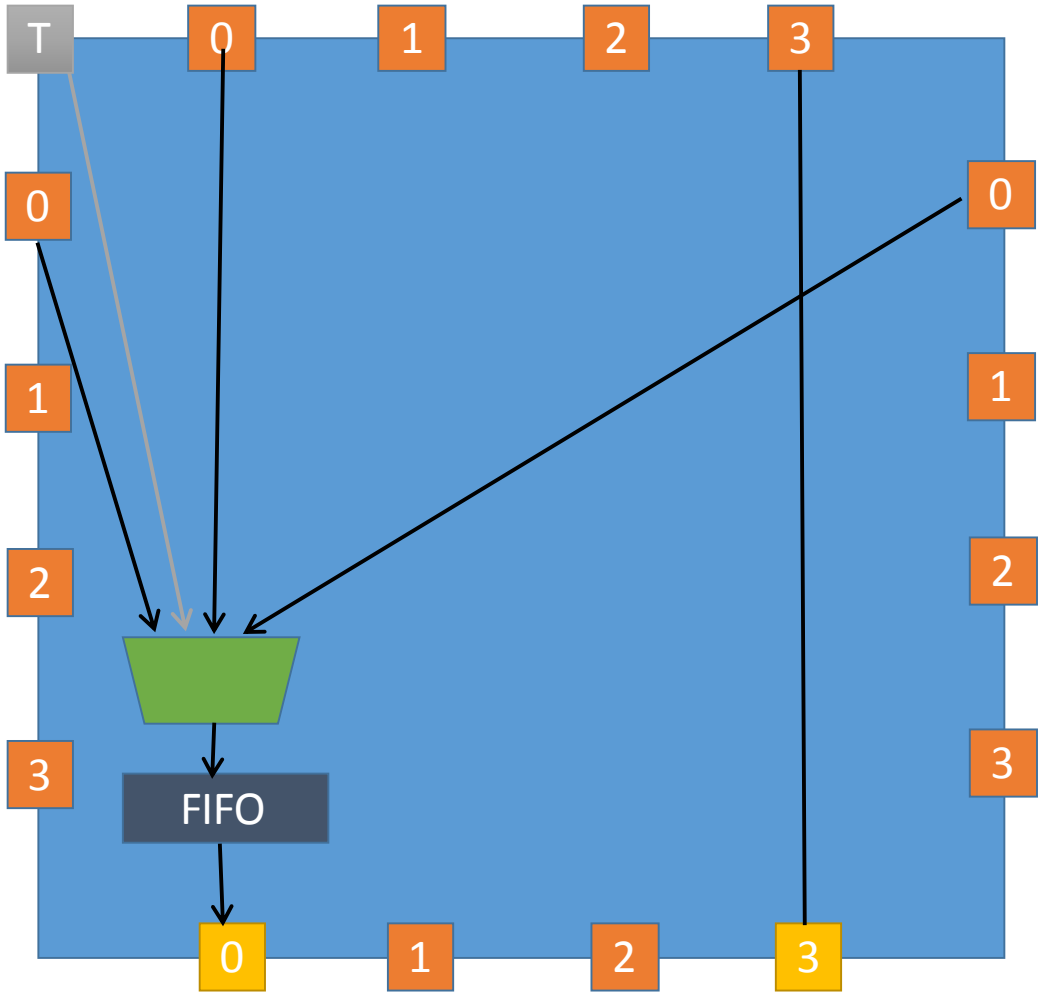
CGRA Architecture



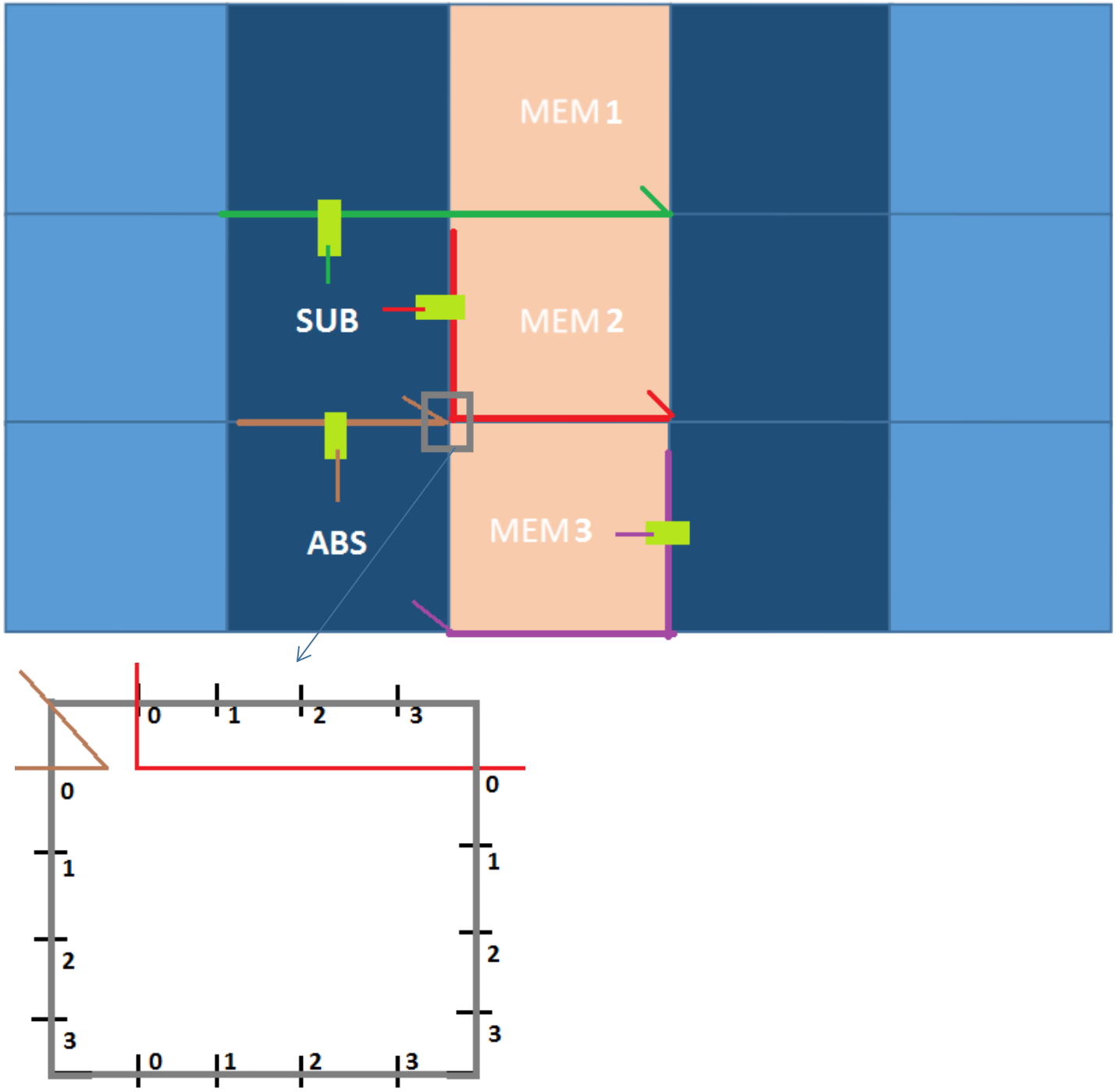
- CGRA fabric consists of a non-homogeneous grid of tiles. The tile pattern is user configurable.
- P 16-bit buses run along tile edges
- Q 1-bit buses run along tile edges
- Each tile has 2M connection boxes that funnel data from the 16 bit buses into the tile (M boxes connect to the buses on the top, while M connect to the buses on the right). Each connection box selects one of P buses.
- Similarly, there are 2N connection boxes that funnel data from the Q 1-bit buses into the tile.

- Each tile consumes 2M input words and performs one or more ops to generate one or more output words
- A switchbox is present at each bus intersection. It routes the buses coming in from a side to any of the other 3 directions. It also injects the output(s) of tile to the top and left of it into any of the buses going out. Switch boxes also have pipeline registers that can delay the incoming data on a bus by a programmable number of cycles.
- Memory tiles contain SRAM modules and associated control. Memory access can be controlled by the application, or it can be a fixed-pattern (eg: linebuffer)



Switch Box

- Switchboxes route buses, delay signals, and inject tile outputs into the routing
- There are P ports on each side of a switchbox that connect to the 16-bit buses
- Similarly, there is a switchbox for the 1 bit signals that is not shown here
- The T port injects the tile output into the switchbox
- Each port can be configured as an input or output port
- Output ports can be driven, or can be pass-through
- Pass through output ports are driven by the input port opposite to them (eg: bus 3 at top must drive bus 3 at the bottom)
- Driven output ports can take their input from any one of the three other sides, or the tile output
- A subset of the driven output ports of a tile can be delayed using a FIFO
- Switchboxes are heterogeneous with respect to the driven/feed-through buses, or FIFO port connections. Eg: Consider four adjacent switch boxes in the horizontal direction with P=4. Bus 3 might be feedthrough for the first three SBs, but is driven for the fourth SB. Thus bus 3 has a segment length of 4 i.e. it can change direction only after 4 tiles. Also, bus 0 might have a FIFO connection in the first SB, but not in SBs two to four.



- This image shows the placement and routing of a trivial C = ABS(A-B) function. A resides in memory mem1, B in mem2, and C is stored in mem3.
- Data from mem1 (green) is injected into one of the P horizontal buses in row 1 by the SB on the lower right corner of mem1.
- Data from M2 (red) is injected on one of the P horizontal buses in row 2. This bus turns a corner in the grey SB and is routed in the corresponding vertical bus
- Green bus can be routed on a segment with length 2, while red bus can use segments of length 1 only (given the placement of the operations)
- The output of the SUB operation (brown) is injected into a bus in row 2 by the grey SB.
- The output of the ABS (purple) operation is similarly routed to mem3, whose CB selects it as the write data for the memory. We could have routed the purple data through the grey SB too.

Questions – Please add to the list

- How does the mapper map and chain line buffers across memory tiles?
- What ops are supported by the tiles?
- What debug widgets are needed?