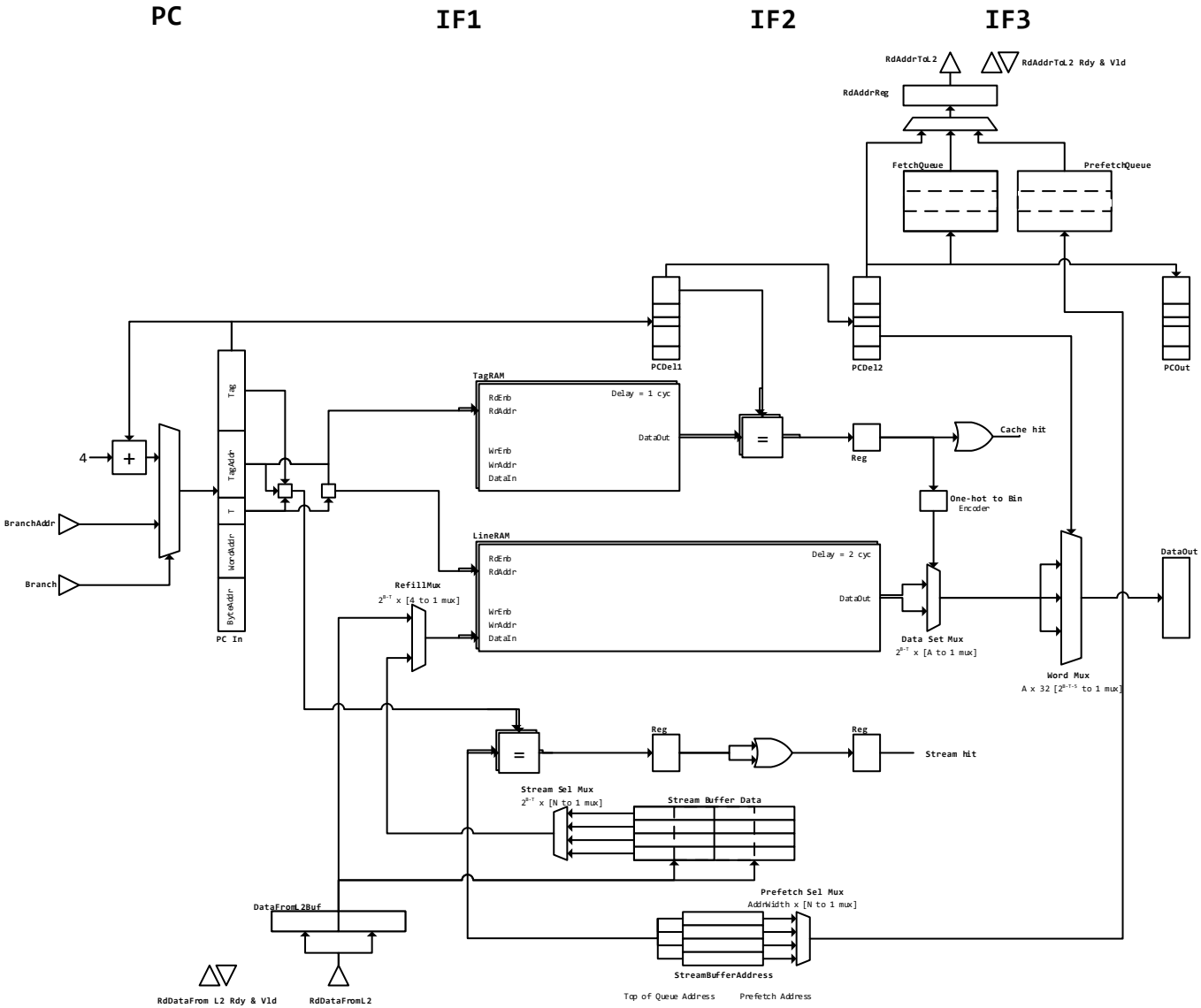

Instruction Cache

- Parameter definitions
  - S – 2^ S is the size of the cache in bits
  - B – 2^B is the number of bits in a cache line
  - T – Each line is stored in 2^T 'sub-sections' in the Line Memory
  - W – 2^W is the width of the data buses towards L2 cache
  - A – Associativity of the cache
  - N – N will be the number of stream buffers
  - n – Each stream buffer will hold 2^n cache lines
  - p – 2^p will be the depth of the prefetch queue

- TagRAM stores the valid bits and tags while LineRAM stores the cache lines
- Instruction cache is a read only cache, with a simple ready-valid interface towards the L2 caches or memory

- LineRAM is allowed two clock cycles to complete a read, while the smaller Tag RAM is allowed only one cycle

- Therefore 3 pipeline stages in total required for a data read or write
  - IF1 – Tag search and line search
  - IF2 – Tag comparison and line search
  - IF3 – Data multiplexing and read/write

- There are multiple stream buffers available which observe the history of instruction fetches and thus predict/prefetch instructions
- Prefetch queue is allocated for storing such requests, though it has access to the bus only with a low priority
- Stream buffers are allocated and deallocated with a LRU policy

- Stream buffers and L1 cache are searched in parallel
- A stream buffer hit will ensure a quick miss recovery, and the hit line will be brought to the L1 cache, and stream buffer will proceed with prefetching further into the pattern
- In case of both L1 and stream buffer miss, a high priority request is sent towards the memory and processor is stalled
- Other two requests in the IF1 and IF2 stages are also immediately checked for misses and requests are sent to L2 if necessary (thus reducing miss penalty for requests in IF1,IF2)

- Critical-word-first fetching and early restart capable